CREATING A

more effective and sustainable housing development model

FOR NORTHERN IRELAND

OCTOBER 2019

TECHNICAL ANNEX AND DATA PROFILE

Data Analysis Methodology

Planning applications data was obtained in spatial format from the Dept of Infrastructure NI for applications received between periods between 28/01/2015 and 28/12/2018, this comprised details of 27134 residential applications.

This data was assessed and categorised into,

- “New Build”,
- “Extensions”,
- “House Type Change” and
- “Holiday/Student Accommodation”.

During this process it was noted that 12,872 of the applications had blank cells in relation to the number of units proposed in the application. This information was obtained from the description details of the application and added into the blank cells. The remaining applications, number of residential units data, was checked to ensure that the number of units applied for, as quoted in the proposal description details, matched the number in the units in the data field. A substantial number of amendments were made to ensure that the data was consistent between both, with the proposal description details information taking primacy. The time taken for applications to be processed in weeks was also calculated from the applications received date to the decision date.

As the focus of this study is on residential applications the New Build proposals were extracted from the data set and a series of spatial analysis and attribute joins were undertaken using GIS software. Spatial data sets were joined to determine if the location of the application sites intersected with any designated policy areas or were within close proximity to environmental factors such as contaminated land or noise sources. These policy designations and environmental considerations were selected as they potentially could increase the level of scrutiny of proposals and might require additional assessments to be undertaken, necessitating additional time to complete the application determination process. Conversely there might also be an expectation that those proposals which fall within designated housing zonings within Local Development Plans on greenfield sites, might have a smoother progression through the application process and therefore be dealt with more expeditiously.

The following data sets were spatially joined to the New Builds data.

a) Settlement Development Limit
b) Local Government District
c) (Built and Natural Heritage Designations) Conservation Area, Area of Townscape Character, Special Area Conservation, Local Landscape Policy Area, ANOB and ASSI.
d) Development Plan Zonings, Housing, Open space, Employment etc.
e) 15m Proximity to Motorway A or B Class Road, re traffic noise impacts
f) Within 60m proximity to a railway line re noise and vibration
g) Proximity to an NIEA identified Brownfield Sites re Land quality issues.
h) Air Quality Management Areas.
i) Potential Waste Water Treatment Constraint
j) Historical Flooding

One completed the joined new build and planning and environmental data was exported to SPSS for further analysis using a proportional ordinal regression model.

Figures 1-4 illustrate the spatial joins that were produced to facilitate the analysis of the data.

Figure 1 New Building Residential Applications with Built and Natural Heritage Designations

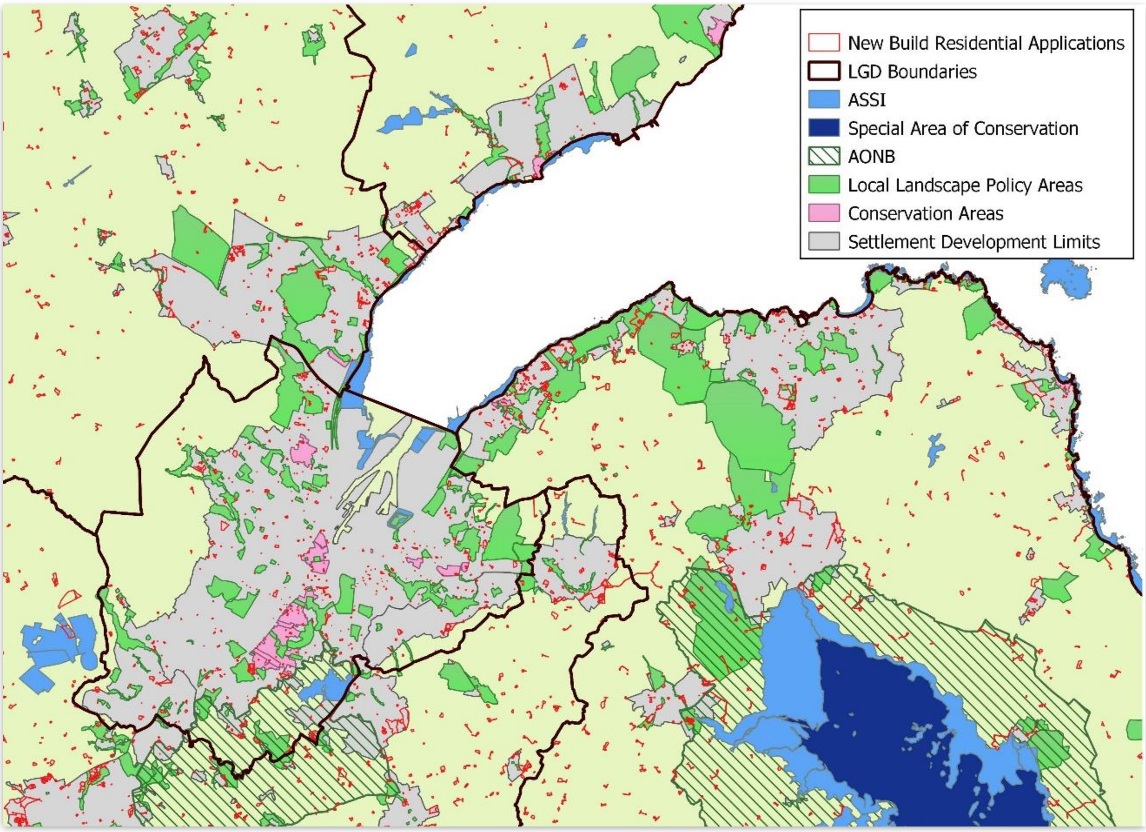


Figure 2 New Building Residential Applications and Local Development Plan Zonings

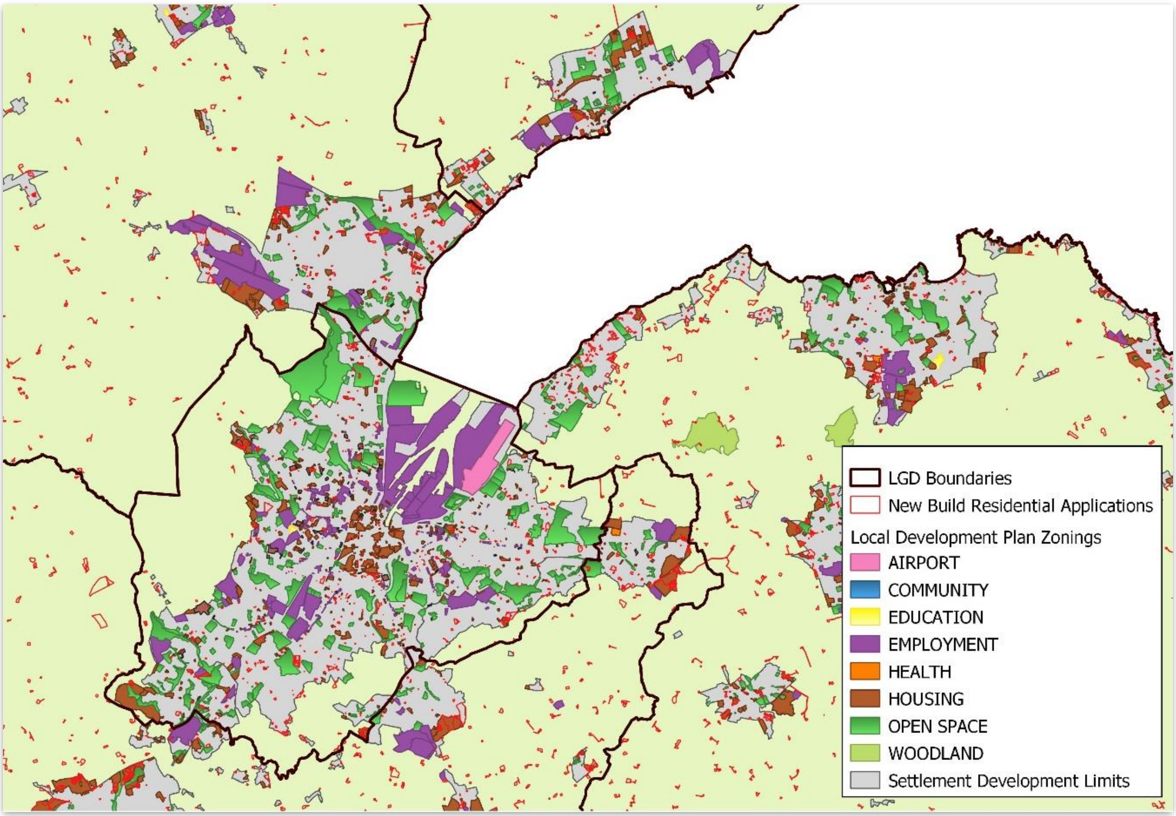


Figure 3 New Building Residential Applications Road and Rail and Air Quality Considerations

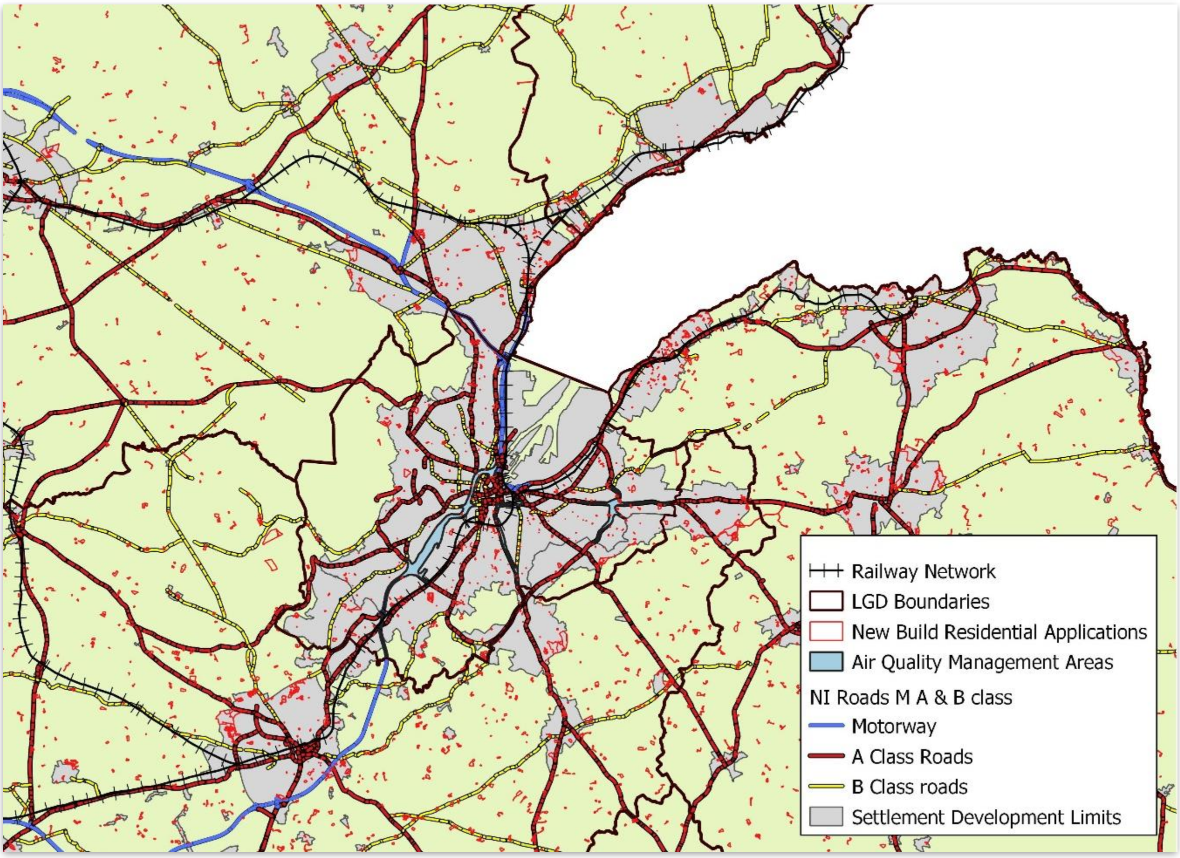
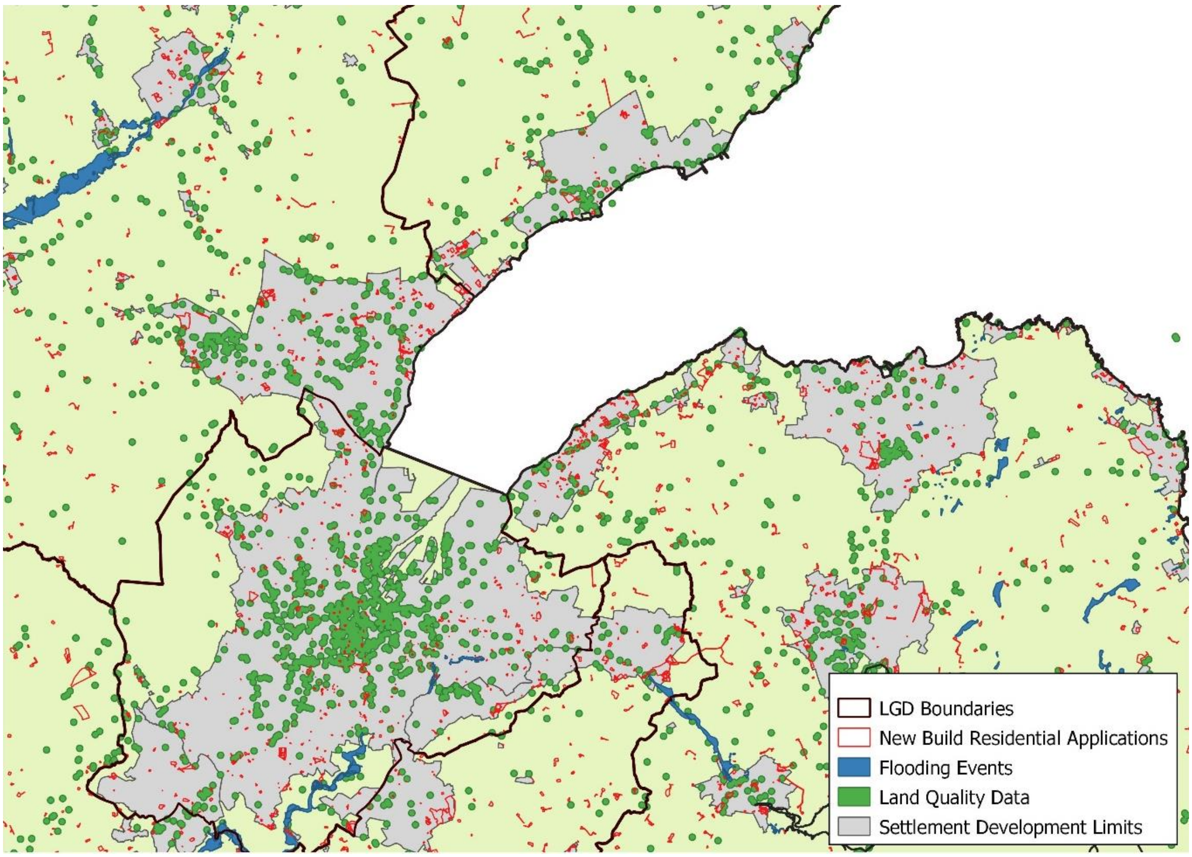
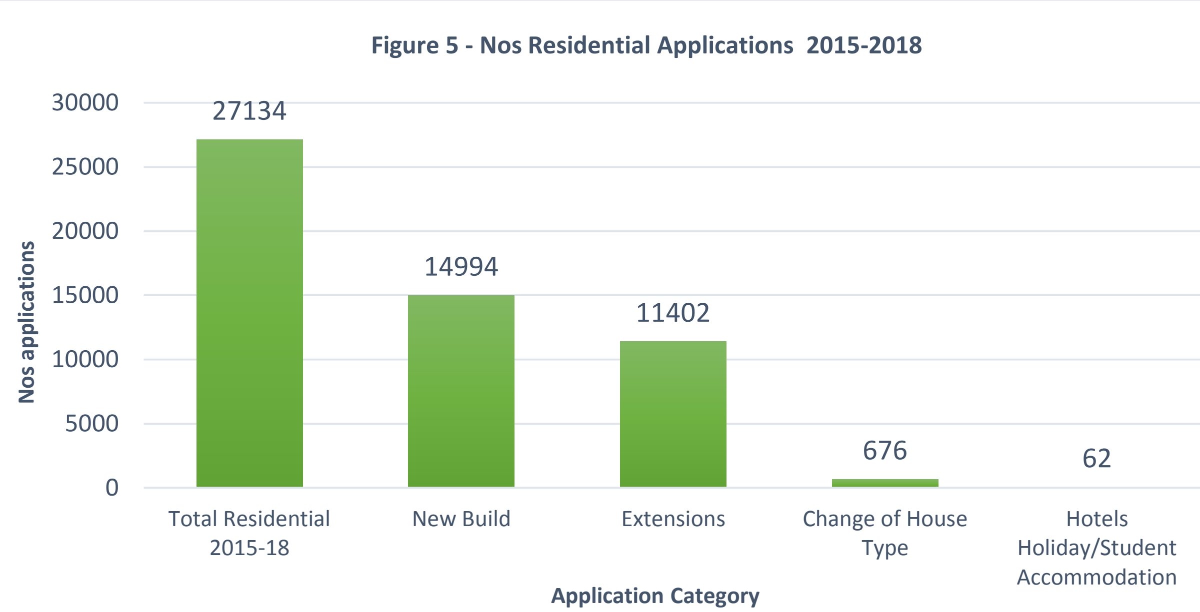


Figure 4 New Building Residential Applications Land Quality and Flooding Considerations

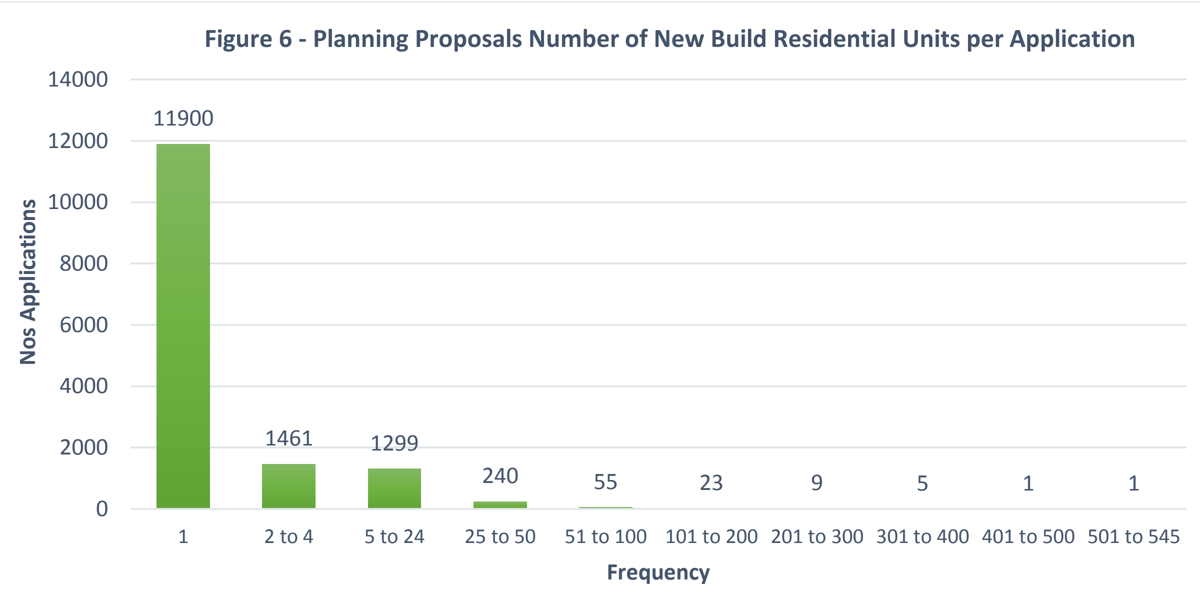


Analysis and Spatial Analysis Findings

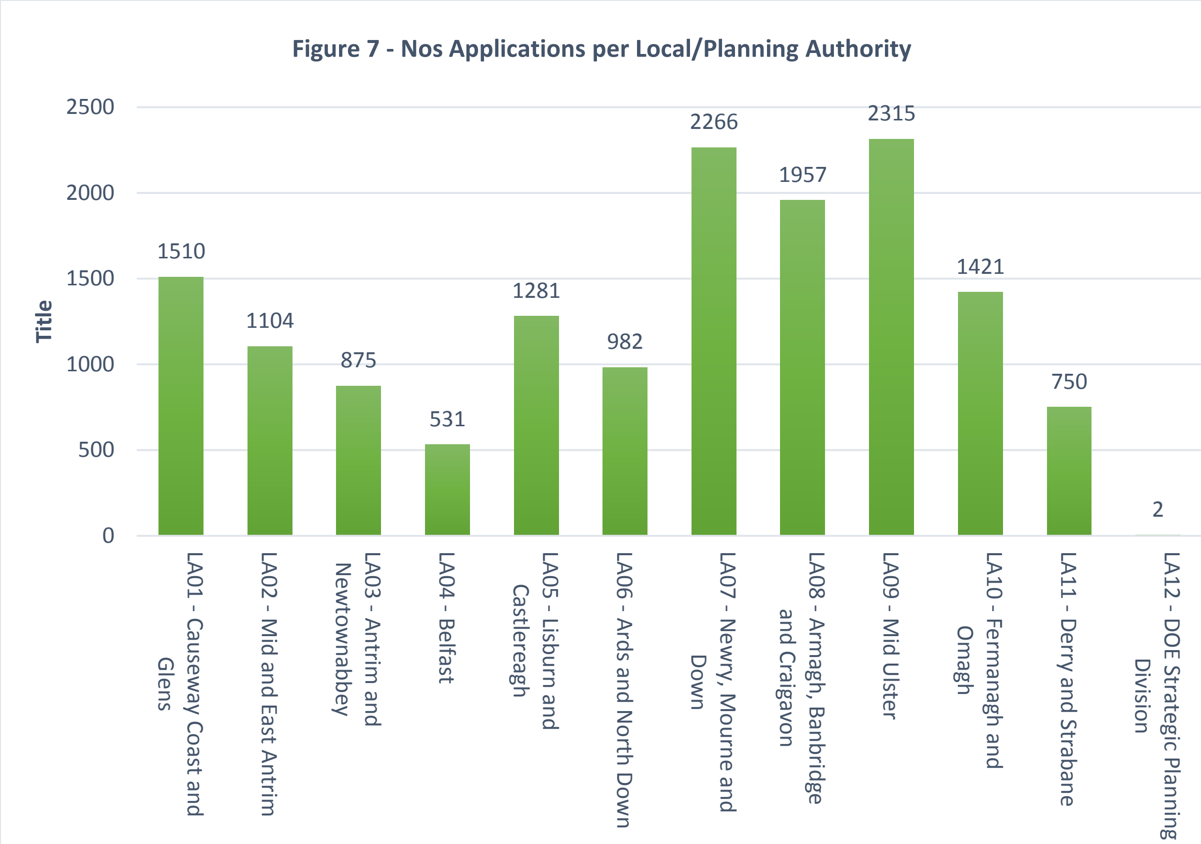
The applications categorisation process into “New Build”, “Extensions”, “House Type Change” and “Holiday/Student Accommodation” produced the following results in Figure 5.



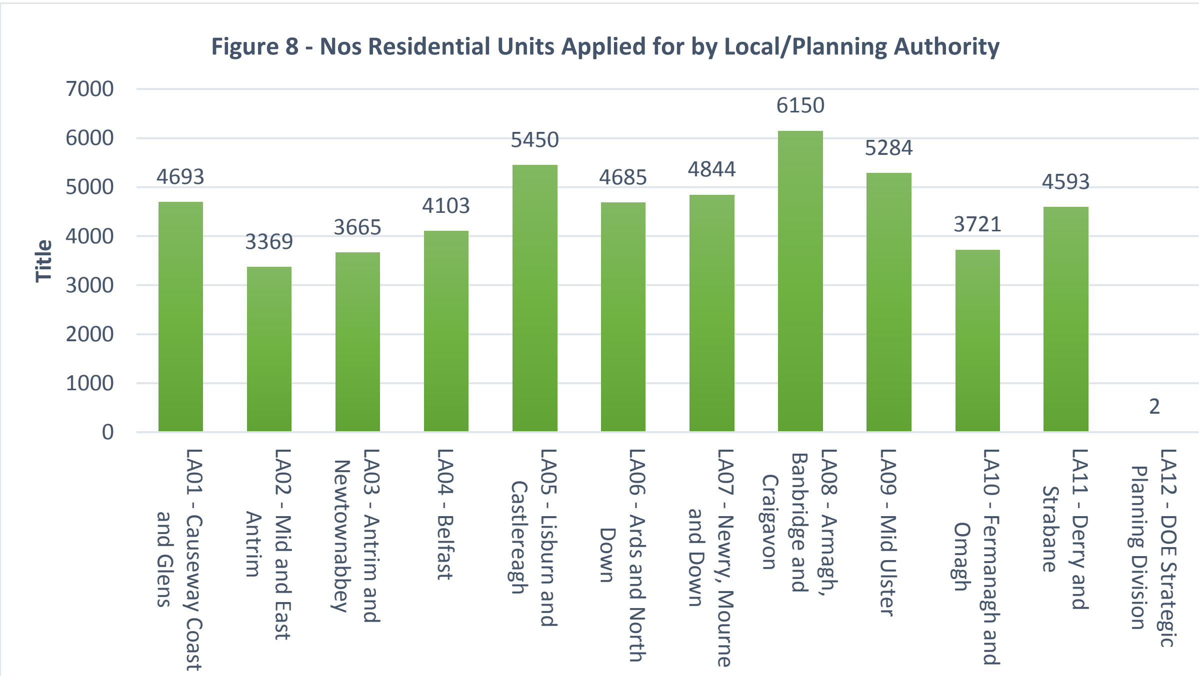
The distribution frequency of the number of residential units per planning application is as follows Figure 6.



As is illustrated by the graph above 79.4% of the new build residential applications are for single dwellings. The number of planning applications by Local/Planning authority is shown in Figure 7.

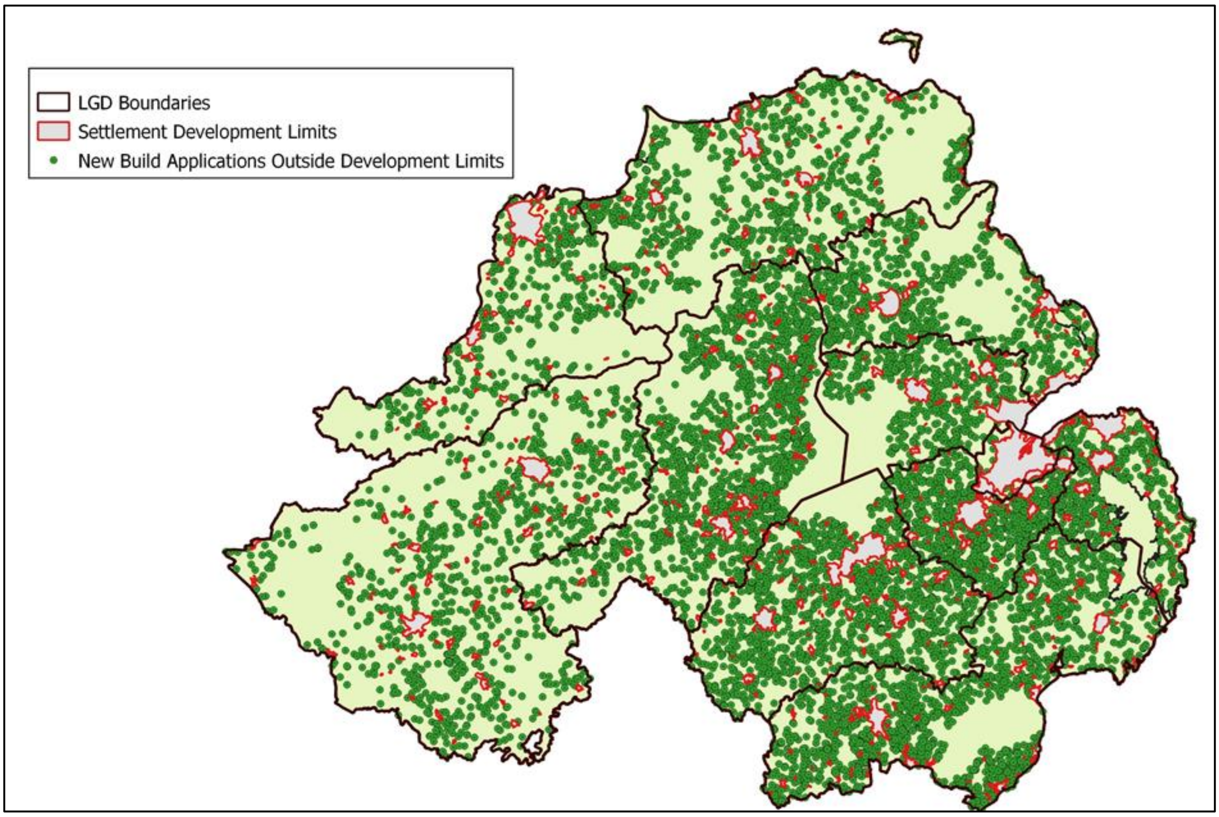


As can be seen the highest number of applications are to be found in Mid Ulster and Newry Mourne and Down Councils. A total number of 50559 residential units were submitted for planning approval during the study period. Figure 8 shows how these were distributed between planning authorities.

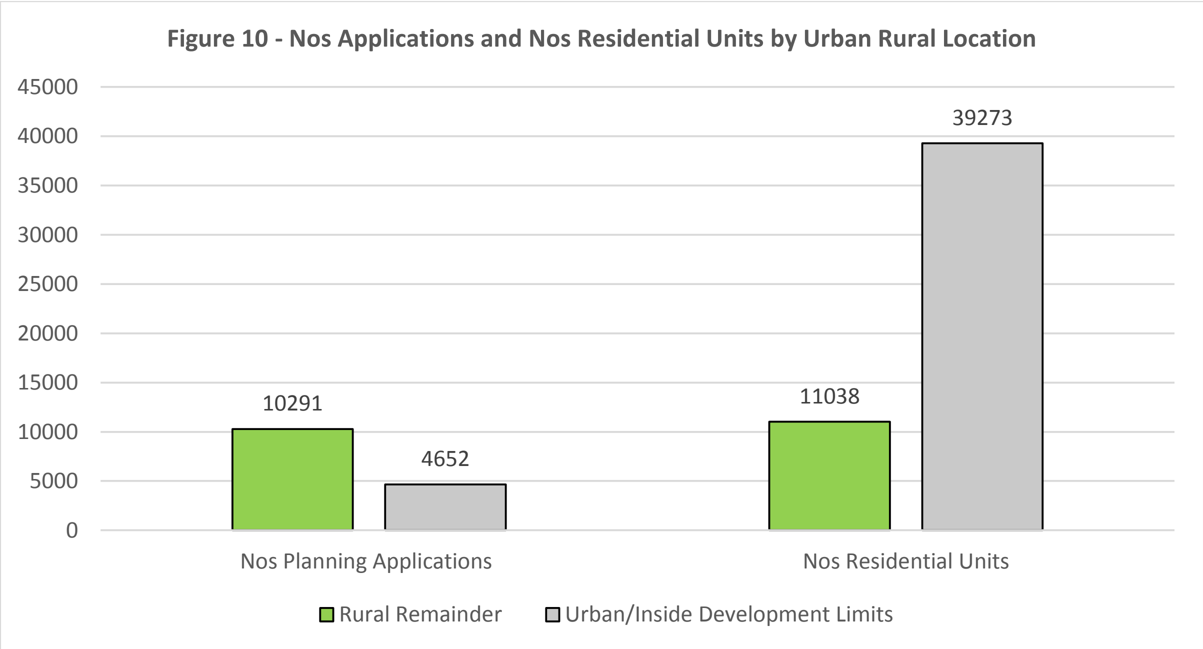


Out of the 14994 applications, 14951 were geo located in relation to their urban/rural location, 4652 or 31.2% were inside settlement development limits indicating that 10291 or 68.8% of application were in greenbelt locations (Figure 9).

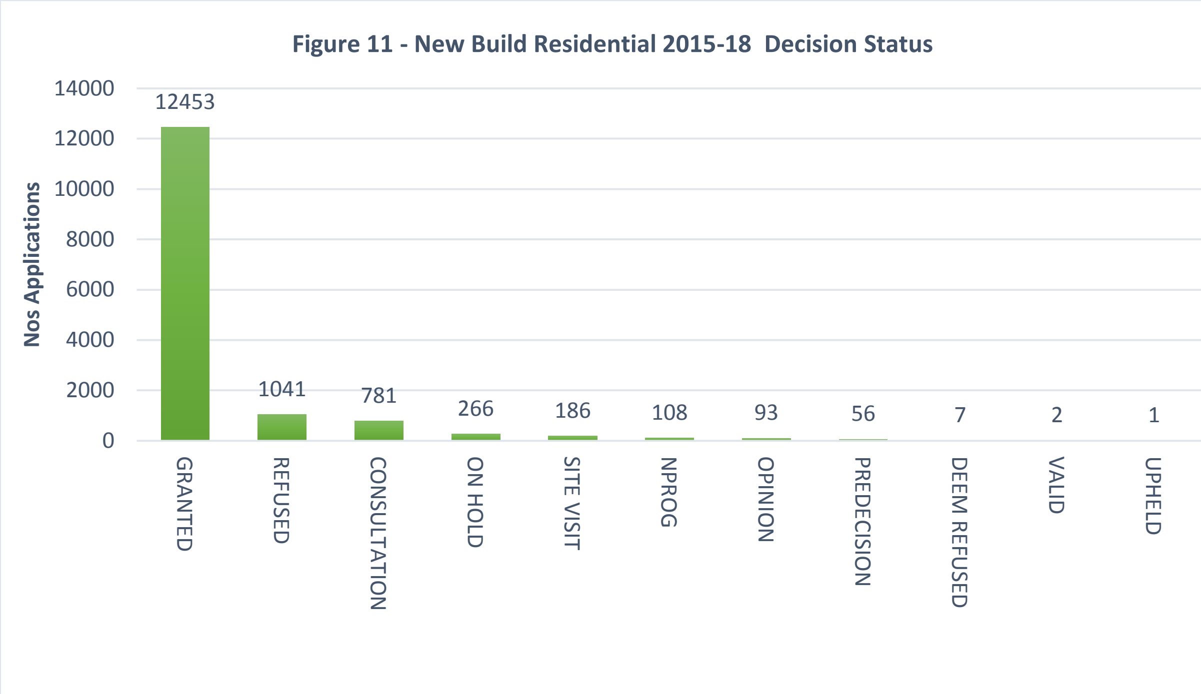
Figure 9 Residential Planning Applications Outside Development Boundaries



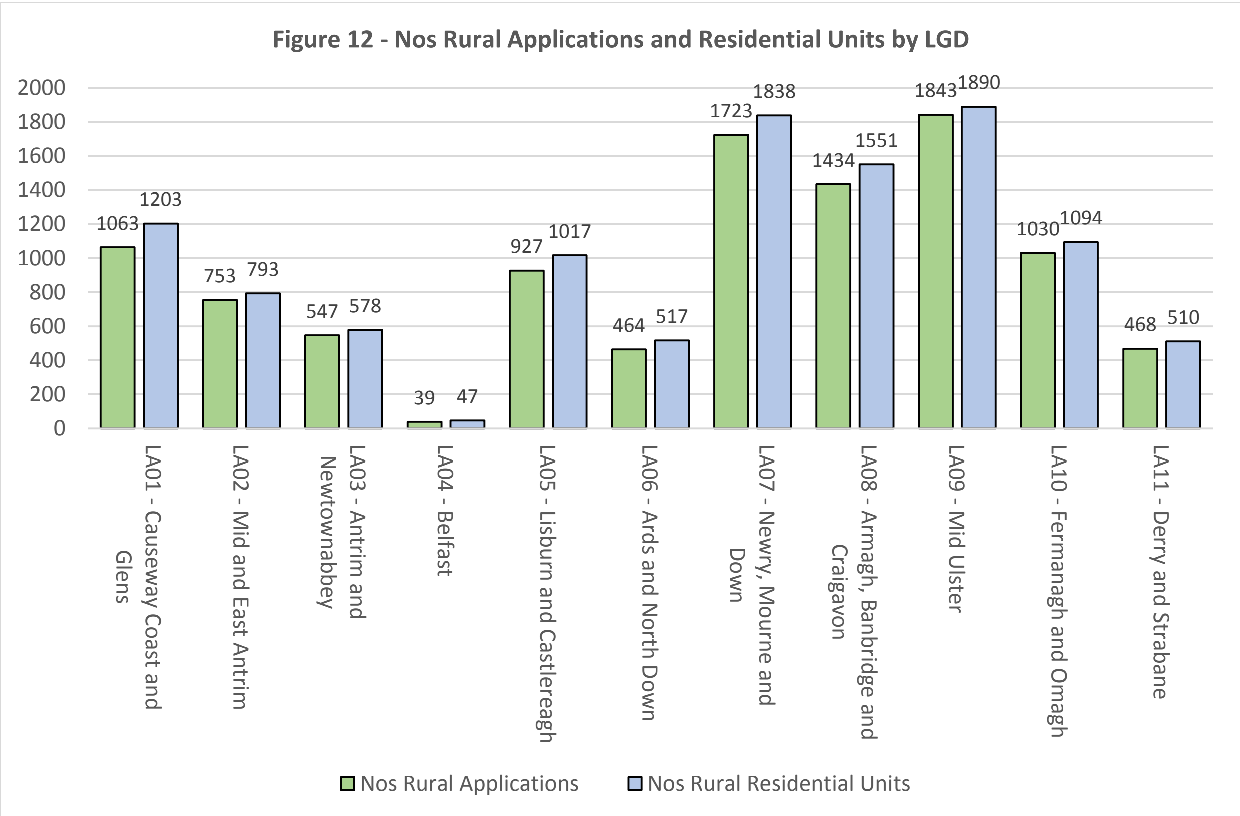
While urban Areas account for 31.2% of the New Build Applications they make up 78.1% of the Nos of Residential Units, conversely Rural Areas account for 68.8% of the New Build Applications but 21.9% of the Nos of Residential Units applied for (Figure 10).



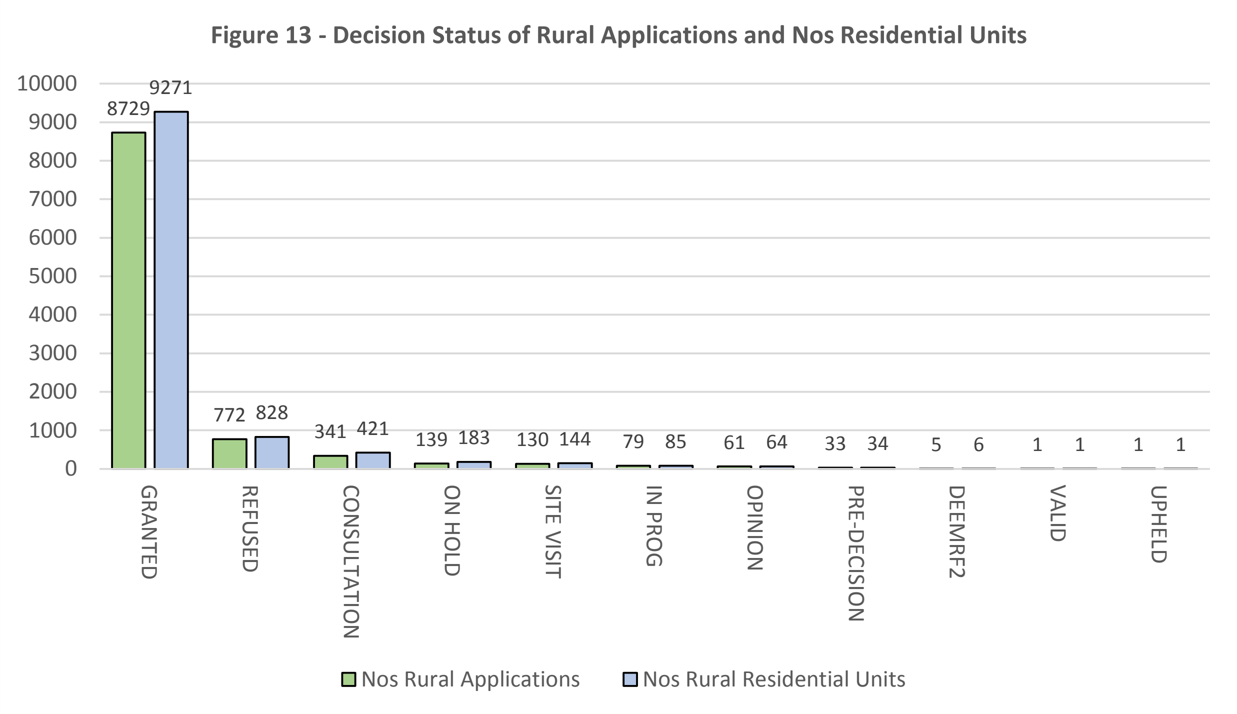
New Build Planning Decisions: of the 14,994 applications made 12,453 or 83% were granted planning permission, while 7% were refused permission (Figure 11). The remaining 10% of applications in the database were in the determination phase i.e. undergoing consultation, awaiting site visits, on hold etc.



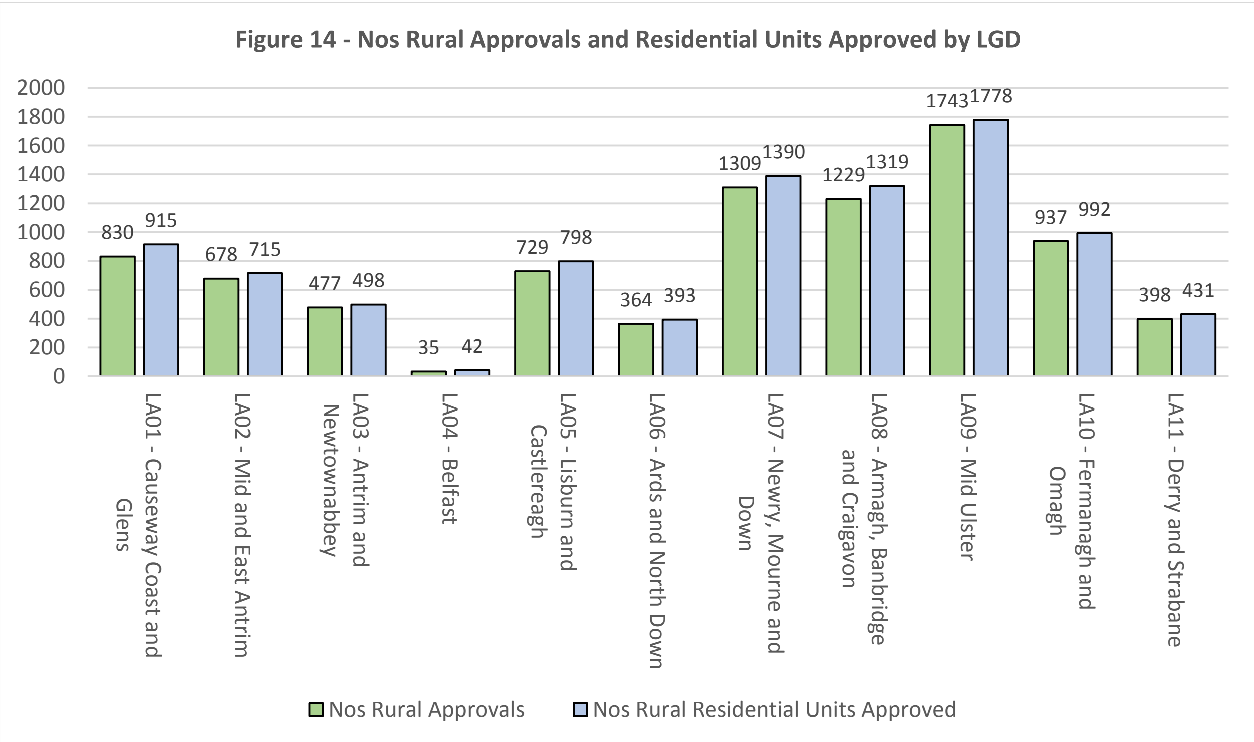
A total of 10,291 applications were made for proposals in the rural remainder. Figure 12 shows the breakdown of the number of applications made and number of residential units applied for by Local Authority.



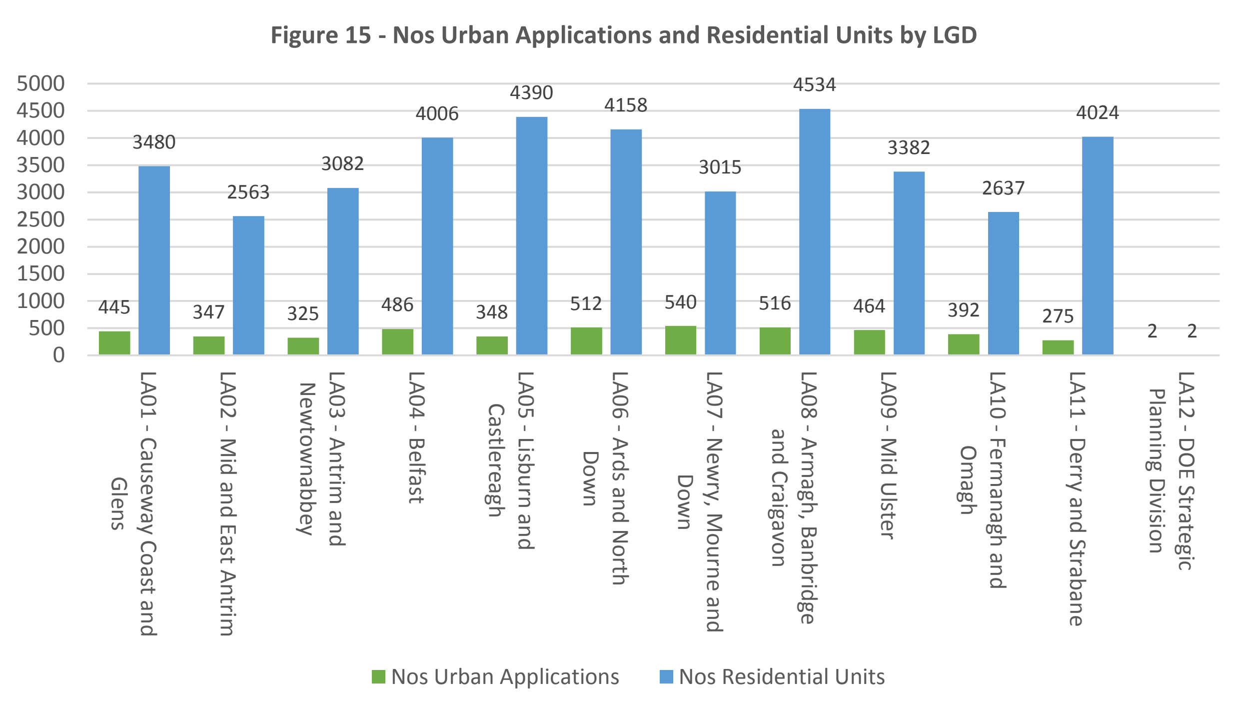
Of the 10,291 Rural Applications 8,729 of these were approved, 772 refused and the remainder were mostly still in the determination phase of the planning decision making process. The approvals accounted for 9,271 residential units in the rural area (Figure 13).



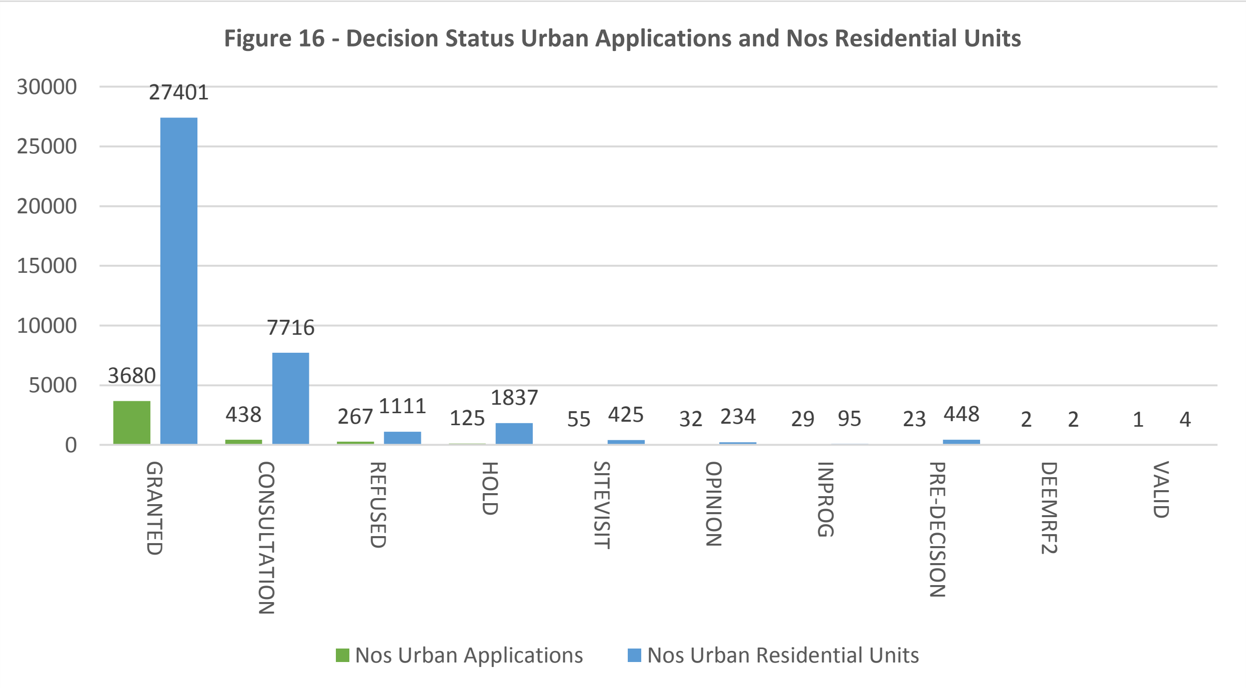
An analysis of the rural approvals and the number of residential units approved in the rural area is by Local Authority is provided in Figure 14.



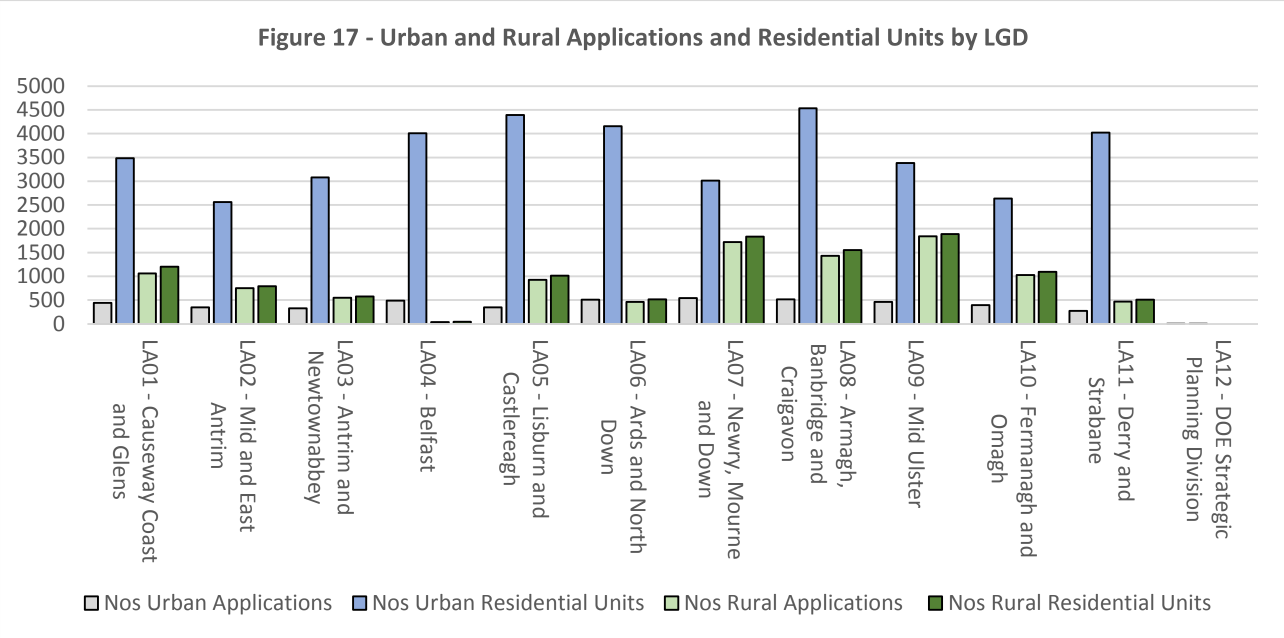
In relation to applications for New Build housing inside settlement development limits the graph below illustrates the distribution across Local Authorities. There was a total of 4,652 applications made comprising 39.273 residential units (Figure 15).



Of the 4,652 Urban Applications 3,680 of these were approved, 267 refused and the remainder were mostly still in the determination phase of the planning decision making process. The approvals accounted for 27,401 residential units in the rural area (Figure 16).



An analysis of the rural approvals and the number of residential units approved in the urban area is by Local Authority is provided in Figure 17.



Descriptives

Descriptive statistics for Application decision days for unit density

	N	*Min*	*Max*	*Mean*	*Std. Dev.*
50+	95	114	1,226	426.71	245.157
25-49	161	39	1,008	372.41	176.514
5-24	1,003	33	1,413	323.42	222.326
1-4	12,371	4	1,358	173.09	136.940

Year	Unit density	N	Min	Max	Mean	Std. Dev.
2015	1-4	2509	4	1358	215.82	167.998
	5-24	246	44	1413	412.78	281.869
	25-49	34	125	987	387.38	214.993
	50+	15	205	1226	593.00	334.770
2016	1-4	3523	20	1182	177.86	146.882
	5-24	339	33	1094	346.54	217.343
	25-49	52	129	1008	393.13	177.254
	50+	25	161	1008	401.12	218.825
2017	1-4	3613	25	760	172.61	128.428
	5-24	271	34	794	294.58	162.749
	25-49	56	84	723	390.04	154.153
	50+	21	173	713	428.05	169.089
2018	1-4	2726	25	456	128.23	75.358
	5-24	147	35	459	173.72	93.599
	25-49	19	39	364	236.95	92.634
	50+	9	114	306	217.56	63.327

LGD	Unit de	N	Min	Max	Mean	Std. Dev.
Causeway Coast and Glens	1-4	1190	29	1161	227.24	137.385
	5-24	90	55	1127	351.31	218.473
	25-49	7	259	730	503.00	191.073
	50+	2	668	1008	838.00	240.416
Mid and East Antrim	1-4	980	20	926	100.00	88.483
	5-24	74	33	1055	241.49	189.913
	25-49	16	125	554	287.13	142.741
	50+	3	183	423	280.00	126.440
Antrim and Newtownabbey	1-4	730	27	548	127.76	78.331
	5-24	92	35	709	251.03	169.435
	25-49	18	119	690	325.94	134.560
	50+	7	207	819	383.86	206.183
Belfast	1-4	314	4	832	216.29	159.186
	5-24	124	50	1097	368.86	234.704
	25-49	18	141	589	367.83	154.230
	50+	11	205	1226	514.73	343.455
Lisburn and Castlereagh	1-4	992	31	1182	230.27	161.807
	5-24	77	74	930	365.90	222.742
	25-49	21	175	875	388.29	189.681
	50+	7	197	904	468.14	237.654
Ards and North Down	1-4	722	46	1237	247.83	167.092
	5-24	90	86	1413	361.07	219.268
	25-49	12	232	793	423.42	160.181
	50+	12	203	986	471.00	234.706
Newry, Mourne and Down	1-4	1926	27	1177	200.49	149.617
	5-24	83	55	1323	356.57	252.421
	25-49	6	292	987	473.83	261.768
	50+	1	215	215	215.00	
Armagh, Banbridge and Craigavon	1-4	1614	36	842	150.66	106.344
	5-24	134	42	1230	304.02	233.487
	25-49	20	39	1008	366.15	219.210
	50+	6	268	809	469.50	214.254
Mid Ulster	1-4	2075	28	1297	148.37	115.294
	5-24	91	43	1271	342.15	240.314
	25-49	19	144	712	436.63	170.516
	50+	2	265	806	535.50	382.545
Fermanagh and Omagh	1-4	1272	21	760	108.85	81.444
	5-24	91	34	783	272.19	178.277
	25-49	17	84	578	362.65	141.628
	50+	7	114	369	195.86	82.291
Derry and Strabane	1-4	556	33	1358	231.57	176.474
	5-24	57	49	904	336.09	218.870
	25-49	7	149	286	213.29	50.490
	50+	12	161	698	383.50	150.694

Supplementary technical report: Inferential modelling of the data

Planning data modelling exercises and methodologies

Standard regression analysis is initially employed to investigate the predictive relationships evident within a planning application and the associated time it takes for processing. This type of modelling can comprise many formats. Indeed, as standard economic theory does not suggest an appropriate functional form to be used in regression analysis, and in the absence of clear guidance, it is appropriate to test several functional forms and utilize a multiple regression equation. Cropper et al. (1988) found that simpler functional forms are superlative for unobserved attributes/characteristics unearthed by the researcher or measured with error and employ a semi-log functional form. In this regard, this research employs both the standard OLS fixed effects linear and natural log ($\log_n$). The Multiple Regression equation takes the form:

$$Y = a_o + b_1x_1 + b_2x_2 + \ \ldots + b_nx_n + \varepsilon$$

Where; a_o - *Is the Regression Constant;* $b_1 \ldots b_n$ - *Are the Regression Coefficients; and* ε *is the Error term.* The basic objective of multiple regression analysis is to develop a strong predictive relationship between property characteristics and value, so that the latter can be estimated through knowledge of the former.

The semi-log linear fit is applied within the modelling frameworks due to computational efficiency and interpretability which provides useful interpretations of the independent variable coefficients in terms of their elasticity in respect to the dependent variable. The semi-log specification is as follows:

$$LnY = \beta_0 + \beta_1.X_1 + \beta_2.X_2 \ldots\ldots \beta_n.X_n + \varepsilon$$

Where; *LnY* is the dependent variable (log of sale price), $X_1 \ldots\ldots X_n$ are the independent variables; $B_0 \ldots \beta_n$ are parameters to be estimated; with ε the error term.

To evaluate the percentage effect, a variation of the equation suggested by Halvorsen and Palmquist (1980) for the semi-log model specification is applied. They point out that unlike a continuous variable, the coefficient of a dummy variable, multiplied by 100, does not represent the usual percentage effect of that variable on the dependent variable. Transformation of the equation applying equation 4 captures the true percentage change:

$$[1 - \ e^{bn}]$$

The estimated true percentage effect of a dummy variable is therefore equal to:

$$100\left(e^{bn} - 1\right) \text{ or g = exp([}\alpha\text{]) -1,}$$

Where; the relative effect on the dependent variable of the presence of the factor represented by the dummy variable bn.

Binary Logistic Regression

Within this research, the dependent variable is transformed into a dichotomous state therefore requires the generation of models for predictions based on likelihood of a planning application being determined within either a 15 week period or a 30 week period (i.e. to predict measuring variables for the probability of whether an application is determined or not within these time periods based on the characteristics). When categorical, the assumption on linearity is violated and logistic regression can be used to transform the linear model in logarithmic terms (*logit*) permitting the prediction of categorical outcomes based on the probability of occurrence. Instead of predicting the value of *Y* from a predictor variable(s) $X_{(n)}$ we examine the dichotomous prediction of probability of *Y* occurring (P)*Y* from known values (e = natural logarithms) resulting in probability of *Y* occurring equating to the case belonging to a particular category culminating in a binary estimation (0; 1).

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1X_{1i})}} \text{ or } P(Y) = \frac{1}{1+e^{-(b_0+b_1X_{1i}+b_2X_{2i}\dots b_nX_{ni})}}$$

A value close to 0 suggests that *Y* is very unlikely to have occurred, with a value close to 1 implying that *Y* is very likely to have occurred employing a maximum-likelihood estimation procedure which selects the coefficients (*β*) that make the observed values most likely to have occurred - in essence, the chosen estimates of the *β*s will be ones that, when values of the predictor variables are placed in it, result in values of *Y* closet to the observed values. Assessing the model, *the log-likelihood*, is based on summation of the probabilities associated with the predicted, P(Y_i) and actual Y_i, outcomes – similar to residual sum of squares (RSS):

$$\sum_{i=1}^{N}[Y_i In(P(Y_i)) + (1-Y_i)In\left(1-P(Y_i)\right)]$$

The model is assessed using the likelihood ratio, illustrating that a negative coefficient value implies that as a predictor value increases, the likelihood of the outcome decreases, with a positive value indicating that as the predictor variable increases, so does the likelihood of the event occurring (Field, 2018). The predictors are assessed within the model by examining the individual 'fit' employing the Wald statistic (*z*) and odds ratio (Exponential of *β*). The *z* statistic[1] indicates whether the *b*-value for the predictor is significantly different from 0; illustrating its significant contribution to the prediction of the outcome (*Y*). The odds ratio reflects the exponential of *β* and is an indicator of the change in odds resulting from a unit change in the predictor, with the odds of an event occurring defined as the probability of an event occurring divided by the probability of the event not occurring:

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1X_{1i}+b_2X_{2i}\dots b_nX_{ni})}}$$

Where the Odds:

$$= \frac{P\ (event)}{P\ (no\ event)};\ P(event\ Y) = \ \frac{1}{1+e^{-(b_0+b_1X_{1i}}}\ ;\ P(event\ Y) = 1 - P(event\ Y)$$

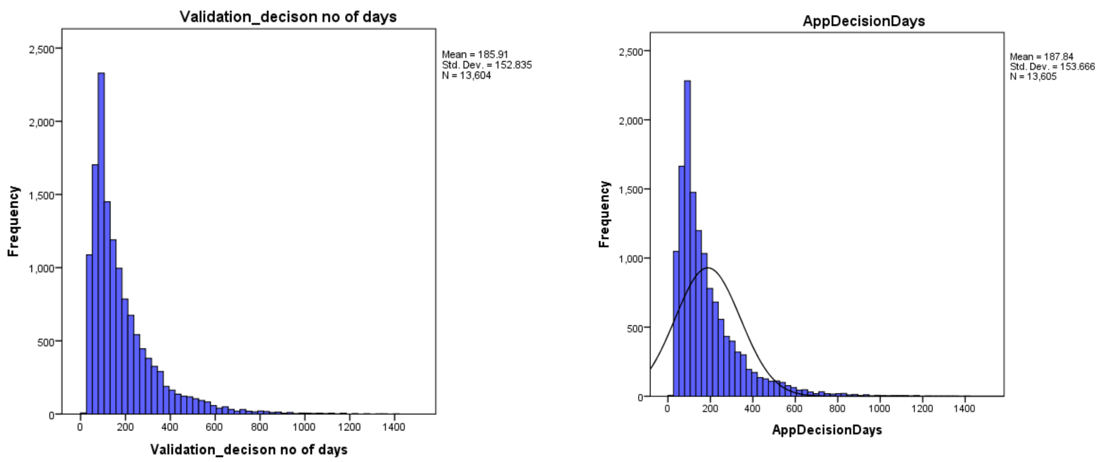
This provides the odds before and after a unit change in the predictor variable, thereby demonstrating the proportionate change in odds (Odds ratio) which can be interpreted as a value exceeding 1 (>1) to show that as a predictor increases, the odds of the outcome occurring increase, with <1 indicating that as a predictor increases, the odds of the outcome occurring decrease.

[1] The Wald statistic is the z^2 Chi-Squared distribution.

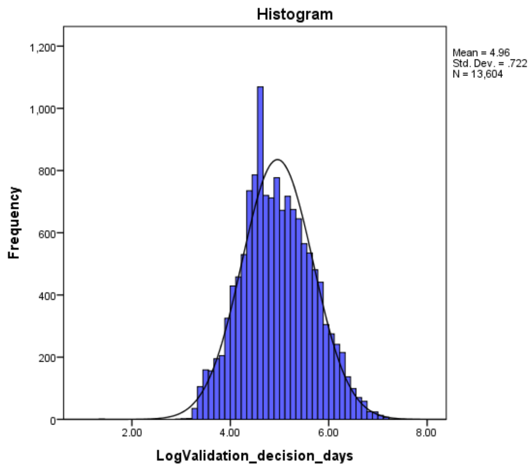
Data

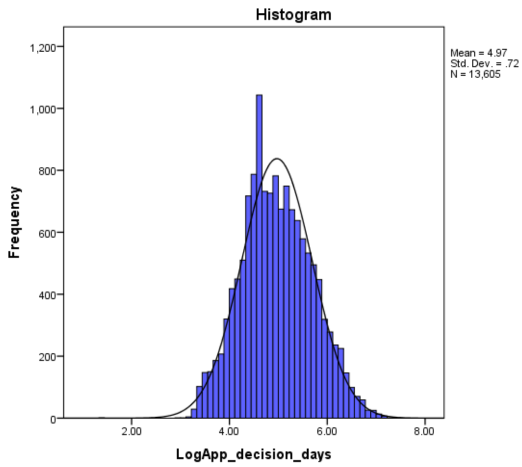
The planning data comprises a number of attributes which can be inferentially tested for statistically significant effects within classical regression hypothesis testing. We have, where necessary, transformed the variables into binary (dummy) format in order to test the inclusion or absence of a categorical effect that may be expected to shift the outcome (Kleinbaum et al., 1988)[2]. This transformation was also conducted at the lowest viable level applicable for statistical modelling. For example, we have created binary variables for 19 regional flood zones boundaries, and unilaterally tested these within the model framework estimation procedure. However, for robust estimation purposes, this becomes too granular for model validity and reliability, therefore a binary was created to symbolise whether the application received falls within a flood zone boundary (or not). Similarly, for Area of Outstanding Natural Beauty (AOBN) designations, whilst there are 9 categories in the data – for modelling purposes this was collapsed into one binary showing an AONB to be present or not. A number of control variables are also included within the models, such as planning decision/status, to account for any partial differential effects, omitted viable bias or mis-attribution effects. The regression model interpretation is pitched against the hold-out (most frequently occurring observation) which is used as the basis for comparison. For example, for the number of units (unit density coefficient) the hold-out is the 1-24 category (binary), thus we denote any statistically significant coefficient values relative to this category and not against zero. Scrutiny of the decision days variable also exhibits a positive skewness (Figure 18). Accordingly, the logarithmic for both the validation data and received date of application to determination is generated for conformance with statistical properties, assumptions and reliability of results.

Figure 18: Number of day frequencies



[2] a dummy explanatory variable with a value of zero will comprise no influence on the dependent variable, whilst a value of one results in the coefficient influencing the intercept.





Log-linear and Logistic regression modelling

As discerned, the planning data comprises a number of attributes which can be inferentially tested for statistically significant effects within classical regression hypothesis testing. To investigate whether statistically significant determinants of planning delays exist, a series of models are undertaken to ascertain which attributes comprise an effect, and the likelihood of a parameter estimate meeting the 15 week or 30 week determination thresholds respectively. The modelling examines the nature of the decision days initially at the overall NI level, with further regional level models based on Local Government Districts and temporal models examined in order to investigate significant determinants for each regional (Local) area and across each year of the data series respectively to establish which characteristics, if any, may be impacting upon the application processing time. We initially test the time taken (number of days) to realise application award from both received date to decision date, and validation date to decision date. The results show that the 'validation' model, which measures the number of days from validating a planning application to the decision date, and the 'received' model (measuring the time taken from application submission to decision date), both display similar findings in terms of coefficient magnitude, direction and significance. In addition, we have also included an additional parameter (The App. Received/validation days) which is a measure of the time taken (number of days) from the date of receipt of an application to when it is validated for inclusion into the model architecture to control for this initial processing time. Further scrutiny of this aspect reveals that 12,151 (81.4%) of applications comprise the same date of receipt and validation, with 94.7% being validated within 10 days and 98.3% within 21 days.

NI Level (Global) models

The findings from the NI level model, which includes the entirety of the data series (14,936 observations), can be evidenced in the summary Table A1[3]. The level of explanation, which

[3] Both the findings of the full linear and log-linear models are available upon request. The log-linear equation takes the form: $Ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots \ldots . . \beta_n X_n + \epsilon$

measures the variation of the explanation of the included (independent) variables (parameters) on the dependent variable (number of decision days), across all the NI level models reveal relatively low R^2 and Adjusted R^2 values ranging between 13.4%-15.5%. This relatively low level of explanation illustrates that the planning data, and particularly the characteristics available within the data, do not seemingly capture explaining the length of time it takes to process applications within the planning system – as 85% is left unexplained and observed as random error. In essence, the attributes within the data comprise limited levels of explanation for understanding the length of time it takes to process an application to determination, highlighting that characteristics not accounted for within the data (exogenous factors) are having the largest and most significant effect(s) on processing time.

Whilst suffering from low levels of explanation, it is noteworthy that most of the regression models are significant (p<.05), and there are a number of determinants which comprise a statistically significant relationships. A number of control variables are included within the models, including application characteristics such as planning decision [PAC] and status, settlements size category, LGDs and time as binary variables, to account for any partial or differential effects, omitted viable bias or mis-attribution effects. The inclusion of these characteristics[4] does also signal some interesting discussion points. For example, in terms of settlement bands, these all show statistical significance illustrating that there is no spatial (level) differentiation in terms of impacting upon processing time, reflecting marginal similarities/differences in levels of magnitude across a number of the band categories. [Within the logistic analysis, the odds ratio also show increased or decreased likelihood for settlement bands to impact upon application time]. This implies that there is no common or typical scale differentiation spanning the rural, peri-urban to urban classifications for application prioritisation – seemingly they are all treated equally. The regression model interpretation is also pitched against the hold-out category (most frequently occurring observation) which is used as the basis for comparison. For example, for the number of units (unit density coefficient) the hold-out is the 1-24 category (binary), thus we denote any statistically significant coefficient values relative to this category and not against zero.

With regards to model estimation, the number of units, the coefficient is statistically significant at the 1% level indicating that, not unsurprisingly, the number of units per application increases the length of decision days. Further examination of this relationship was undertaken using correlation analysis which reveals a statistically significant (p<.001) weak positive association of 21.8%, meaning that the relationship between the increase in time taken to process an application and the increase in unit density (number of units per application) display a relationship of 21.8%. In addition, further banding the units per application to account for unit density categories (1-4 units; 5-24 units; 25-49 units; 50+ units) exhibit these binary variables to also be significant parameters[5]. The coefficient estimates show there to be a high effect on decision time for applications with between 5-24 and 25-49

[5] We have, where necessary, transformed the variables into binary (dummy) format in order to test the inclusion or absence of a categorical effect that may be expected to shift the outcome (Kleinbaum et al., 1988). A dummy explanatory variable with a value of zero will comprise no influence on the dependent variable, whilst a value of one results in the coefficient influencing the intercept.

units. This level of effect decreases for 50+ units which show a moderate effect in terms of magnitude. This is likely due to the small sample size (n=95) of these applications within the data which constitute a mere 0.6% of all applications. Moreover, this is also in line with some of the interview evidence which has illustrated that in a number of instances there is a prioritisation for applications comprising larger unit density (50+) to be evaluated.

The findings further infer that if an application is within an AONB designation, Conservation area, ASSI, local landscape policy area or Special conservation area status, there is evidence of significant effects, albeit at varying magnitudes, nonetheless, proximity to railway infrastructure and designated within a flood zone area does not show any significance. In terms of infrastructure, applications within immediate proximity to both A and B roads (within 15 metre proximity) display a significant effect, with areas showing sewage constraints also having a significant effect on the number of decision days to determination.

Table A1 NI level Log-linear model summary outcomes

Parameter	Effect	Significant
App. Received/validated days	Low	Yes
No. of Units in application	Low	Yes
Unit density 5-24	High	Yes
Unit density 25-49	High	Yes
Unit density 50+	Moderate	Yes
Conservation area	Moderate	Yes
Landscape policy area	Low	Yes
Special conservation area [SAC]	Moderate	No
ASSI	High	Yes
A Road	Low	Yes
B Road	Low	Yes
Railway	Low	No
Flood zone	Low	No
Sewage constraint	Low	Yes
AONB	Moderate	Yes
R^2	0.144	
Adj. R^2	0.141	
F stat.	58.423***	
N	14,936	

NB. Low<.1; Moderate<.5; High>.5%. *denotes statistically significant.
Parenthesis denotes a negative effect

Regional level models

The regression analysis is further undertaken to produce regional level models. The rationale for this is to examine whether the parameter effects are uniform across each respective LGD or indeed if there are more localised challenges in each Council area pertaining to the application processing time period based on the data (Table A2). This is envisaged to provide some inferences as to the more unique or idiosyncratic dynamics seen to be causing delays. Similar to the NI level model, the R^2 statistics are low, nonetheless do show increased explanation for the Mid and East Antrim (21.9%), Antrim and Newtownabbey LGDs (24.7%), Mid-Ulster (18.9%) and Fermanagh and Omagh (22.5%) respectively. The number of units within an application shows between 0.4%-1.6% effects per unitary change across a majority of the LGD regions. These however are not statistically significant within all regions with the Causeway Coast & Glens, Mid and East Antrim, Belfast, Fermanagh and Omagh and Derry and Strabane all insignificant inferring that in these LGDs the processing time for determination and number of units are spurious.

With regards to unit density classifications, there is a relatively consistent picture evident, with applications up to 49 units displaying moderate to high statistically significant effects against the base category (1-4 units), illustrating that the changes in unit density do impact upon increased delays. Again, this effect is not significant for Belfast across both the 5-24 and 25-49 unit density categories or Lisburn and Castlereagh and Newry, Mourne and Down at the 25-49 category. Notably, although not significant, both the Fermanagh and Omagh and Derry and Strabane LGDs show moderate negative coefficients. Upon first inspection these seemingly appear counterintutative, this is perhaps more suggestive of more expedient processing time of these larger applications in these regions. An interesting finding emergent within the modelling indicates that applications with 50+ units demonstrate no statistical significance, with the exception of the Fermanagh and Omagh and Newry, Mourne and Down regions – which are notably negative. Indeed, there is a mixed representation spatially of the effects of the largest schemes. The results are suggestive of larger schemes comprising an impact, both positive and negative but cannot be characterised as statistically significant, and where they are they appear to be prioritised. In essence, the size or scale of applications cannot be determined, in any *real* spatial sense to be significant contributors to causing delays in applications reaching determination.

Accounting for the policy environment to establish whether there is any lethargic policy dynamics impacting upon processing determination, there is again some varied effects across LGDs which naturally alters across the urban plain. Both the Derry and Strabane, Newry, Mourne and Down and Antrim and Newtownabbey regions reveal high positive statistically significant effects suggesting that being located within conservation areas in these LGDs is impacting upon processing time. The remaining LGDs whilst displaying moderate or low effects, both negative and positive are insignificant. These effects are also noticeable for the other planning policy parameters included within the modelling exercises. ASSI shows a statistically significant positive effect in the Fermanagh and Omagh, Newry, Mourne and Down, Mid and East Antrim and Antrim and Newtownabbey LGDs. Proximity to existing road infrastructure does also reveal assorted levels of impact statistically within a number of the LGDs indicating that there may be instances where the existing road infrastructure or lack thereof is creating a time delay in application processing. The findings however, do not show this to be rife. In addition, sewage constraints also appear varied and insignificant although it is worth noting that the Belfast LGD shows a low positive statistically significant effect with the Fermanagh and Omagh region demonstrating a high positive significant effect which may be inhibiting application processing time.

Table A2 Summary of LGD model findings

	Causeway Coast & Glens	Mid & East Antrim	Antrim & Newtownabbey	Belfast
	Effect	*Effect*	*Effect*	*Effect*
App. Received-validated days	Low	None	Low	Low*
No. of Units in application	Low	Low	Low *	Low
Unit density 5-24	Moderate*	High*	Moderate*	High
Unit density 25-49	High*	High*	High*	High
Unit density 50+	High	(Moderate)	(Moderate)	High
Conservation area	(Moderate)	Moderate	High*	Moderate
Landscape policy area	(Low)	Moderate	(Low)	(Moderate)
Special conservation SAC	(Moderate)	None	n/a	Low
ASSI	Moderate	High*	High*	Low
Road Proximity A road	Moderate*	Moderate*	Low	Low
Road Proximity B road	Low	Low	Low	Low
Rail Proximity	Moderate	(Moderate)	(Low)	Low
Flood area	Moderate	Moderate	(Moderate)	Low
Sewage constraint	(Moderate)	(Moderate)*	n/a	Low*
AONB	Low	Moderate*	n/a	Low*
R^2	0.1006	0.238	0.269	0.210
Adj. R^2	0.0873	0.219	0.247	0.167
F stat.	4.785***	12.55***	12.112***	4.885***
N				

NB. Low<.1; Moderate<.5; High>.5%. *denotes statistically significant.
Parenthesis denotes a negative effect
n/a. no observation present or is a constant

	Lisburn & Castlereagh	Ards & N. Down	Newry, Mourne & Down	Armagh, Banbridge & Craigavon
	Effect	*Effect*	*Effect*	*Effect*
App. Received-validated days	Low*	Low	Low	Low
No. of Units in application	Low*	Low*	Low*	Low*
Unit density 5-24	High*	High*	High*	High*
Unit density 25-49	Moderate	High*	High	High*
Unit density 50+	(High)	Moderate	(High)*	(Low)
Conservation area	Moderate	(Moderate)	High*	High
Landscape policy area	Low	Moderate	Moderate	(High)*
Special conservation SAC	n/a	Low	(High)	n/a
ASSI	Moderate	(Low)	High*	Moderate
Road Proximity A road	Moderate	Moderate	Moderate*	Low
Road Proximity B road	Low	Low	Low	Moderate*
Rail Proximity	High	Moderate	High*	Moderate
Flood area	n/a	Moderate	(Moderate)	Moderate
Sewage constraint	(Moderate)	Moderate	Moderate	Moderate
AONB	Moderate*	Low	Moderate*	n/a
R^2	0.131	0.136	0.130	.143

Adj. R^2	0.109	0.105	0.117	.129
F stat.	5.768***	4.384***	9.864***	9.711***
N				

NB. Low<.1; Moderate<.5; High>.5%. *denotes statistically significant.
Parenthesis denotes a negative effect
n/a. no observation present or is a constant

	Mid-Ulster	Fermanagh & Omagh	Derry & Strabane
	Effect	*Effect*	*Effect*
App. Received-validated days	Low	(Low)	(Low)*
No. of Units in application	Low*	Low	Low
Unit density 5-24	High*	High*	Moderate
Unit density 25-49	High*	(Moderate)	(Moderate)
Unit density 50+	(High)	(High)*	Moderate
Conservation area	Moderate	(Low)	High*
Landscape policy area	Low	(Low)	n/a
Special conservation SAC	(Moderate)	Moderate	High*
ASSI	(Moderate)	High*	(Moderate)
Road Proximity A road	Low	Low	(Low)
Road Proximity B road	Low*	Low	Moderate*
Rail Proximity	n/a	n/a	n/a
Flood area	n/a	Moderate	Low
Sewage constraint	Low	High*	Low
AONB	Moderate*	Low	Low
R^2	0.199	0.240	0.156
Adj. R^2	0.189	0.225	0.122
F stat.	20.612***	15.885***	4.663***
N			

NB. Low<.1; Moderate<.5; High>.5%. *denotes statistically significant.
Parenthesis denotes a negative effect
n/a. no observation present or is a constant

Time Models

The data is also dissected to produce regression models for each distinctive annual time period over the data series (Table A3). This exercise was undertaken in an attempt to establish whether there appears to be any changes in terms of significant parameters across the different annual time periods, and indeed, if any indicators remain prevalent. The results infer that the time to validate an application from original receipt whilst significant and comprises a low effect (0.1-0.9%). This also equates to the number of units, or unit density, for the applications which remains a significant parameter for each of the four years analysed. When banding the unit density there is slightly more variation in terms of the magnitude of the effects and level of statistical significance. Applications comprising 5-24 units and 25-49 units indicate a much higher effect for the annual periods of 2015, 2016 and 2017, however it is notable that the level of this effect reduces in 2018 and becomes statistically insignificant. For applications with units of 50+, they comprised a moderate statistically insignificant (*p*>.05) effect in both 2015 and 2016. Interestingly, the coefficient for 2018 displays a high negative effect symbolising that these applications with high unit density do not impact upon the time it takes for applications to reach determination. Again, this equates to the view that these applications are being prioritised and also evidenced in the descriptive analysis which exhibits both the average number of days for these 50+ applications to be decreasing and also the volume of them entering into the planning system has diminished over the data series.

The role of conservation status comprises mixed effects, but in the main, are insignificant and comprise marginal effects on increasing application delay with the exception of ASSI designated areas and areas of outstanding natural beauty which show a consistent moderate statistically significant impact between 2015 and 2017 and a low but significant effect in 2018. In terms of trunk infrastructure, there is a varied effect on an annual basis regarding the adjacency to roads and sewage constraints. This appears to have decreased over the four year investigation period, however it is worth noting that sewage constraints as of 2018 comprises a low but statistically significant effect – inferring that these constraints have an impact on the time it takes for application determination.

Table A3: Time period Log-linear model summary outcomes

	2015	*2016*	*2017*	*2018*
	β	*β*	*β*	*β*
App Received/Validated days	Low*	Low*	Low*	Low*
No. of Units in application	Low*	Low*	Low *	Low*
Unit density 5-24	High*	High*	High*	Moderate
Unit density 25-49	High *	High*	High*	Moderate
Unit density 50+	Moderate	Moderate	High*	(High)*
Conservation area	Moderate	Moderate*	(Moderate)	Low
Landscape policy area	Low	Low	Moderate	Low
Special conservation SAC	(Moderate)	Moderate	Low	(Low)
ASSI	Moderate	Moderate	High*	High*
Proximity A road	(Low)	Moderate*	Low*	Low
Proximity B road	Low	Moderate*	(Low)	(Low)
Proximity Motorway	High*	n/a	(High)	n/a
Rail Proximity	Moderate	(Low)	(Low)	(Moderate)
Flood event area	Low	Low	(Low)	Moderate
Sewage constraint	Low	Moderate*	Low	Low*
AONB	Moderate*	Moderate*	Moderate*	Low*
R^2	0.168	0.17	0.149	0.084
Adj. R^2	0.156	0.163	0.142	0.074
F stat.	15.830***	24.266***	19.675***	8.262***
N	2,817	3,973	4,170	3,976

NB. Low<.1; Moderate<.5; High>.5%. *denotes statistically significant.
Parenthesis denotes a negative effect

Logistic models

The model pre-tests shows the simultaneous inclusion of all variables using the omnibus test of model coefficients, a likelihood ratio test. The model Chi-square value (χ^2 =3956.42, *p*<.001). reveals the model to be statistically significant at the 1% level showing the model to improve the explanation above the intercept model, with the Nagelkerke R^2 statistic[6] demonstrating a level of explanation of 32.2%. The Hosmer and Lemeshow Chi-square test further tests the model fit using non-significance as the 'goodness of fit'. The test, akin to original model testing shows the model to be statistically acceptable (χ^2 =14.210, *p*>.05 [0.126]). The model classification table essentially plots the observed group memberships against the predicted relationships within the logistic regression model. The model correspondence between the observed and predicted memberships shows 72.9 percent correctly predicted cases.

In terms of model evaluation, standard least squares regression indicates that the coefficients denote the predicted amount of change on the dependent for every unitary increase on the predictor variable. In context of logistic regression, the dependent variable is dichotomous thus symbolises the predicted probability of falling into a target group, in this case if applications submitted fall within the 15 week processing period. In the logistic model the relationship is non-linear, therefore the conventional practice is to convert the probabilities to log-odds which measure the predicted change in log-odds[7] for every unitary increase in the predictor variable. As observed in the methodology, this equates probability to the probability between events (A; B) creating a ratio of probability where the Odds ratio null hypothesis is equal to 1, reflecting the idea that there are no changing odds as a function of increases or decreases within the predictor variables.

As the logistic model is structured based on the value of 1 equating to meeting the 15 week threshold (coding system; 1 = 15 weeks, 0 = beyond 15 weeks). The negative coefficients for the number of units within each application indicates that there is a statistically significant (*p*<.001) greater likelihood (odds ratio) that applications with the increased number of units or unit density will fall within the beyond 15 week target group. The odds ratio (the multiplicative change in odds for every one unit change on a predictor) reveals the decreasing probability of being in the target group as scores on the predictor increases – for every one unit increase in the number of units contained within an application, the odds of it increasing beyond the 15 week threshold increase by a factor of 0.869 or 13.1%. When also considering the binary unit density variables (binary unit density classifications), the findings show strong negative coefficients for each respectively, signalling that the probability of the target membership (15 weeks) decreases with increases on the predictor and the odds ratio indicates that the applications are 99% more likely to go beyond 15 weeks.

For applications within areas of landscape policy the positive coefficient indicates that there is an increasing probability of the application falling into the target group on the dependent variable, i.e. there is an increasing probability (1.16 times or 16%) that they are evaluated

[6] Scale is between 0-1, Cox and Snell upper bound does not reach 1.
[7] The natural logarithm of probability of membership in a category.

within 15 weeks as a result of changes in the number of applications within a designated area, although not at any conventional statistical level. Conversely, for designated conservation areas, the odds ratio shows a decreasing likelihood of being in the 15 week period, a similar finding for applications which comprise ASSI. The AONB parameter estimate exhibits this to have a decreasing likelihood of falling in to 15 week determination benchmark, with both sewage constraints and flood event areas not reaching statistical significance. Proximity to road infrastructure shows an increasing effect in terms of odds ratio (1.22 and 1.19 times), indicating that applications with accessible road infrastructure are more likely to be processed within the 15 week time period.

Table A4: Logistic regression coefficients for 15 week processing

	β	Wald	Sig.	Exp(β) OR
Number of units in application	-0.140	44.994*	0.000	0.869
Unit density 5-24	-5.016	11.640	0.001	0.007
Unit density 25-49	-4.472	9.986	0.002	0.011
Unit density 50+	-4.934	9.883	0.002	0.007
Conservation area	-0.615	6.205	0.013	0.540
Landscape policy area	0.153	1.012	0.315	1.166
Special conservation area	0.005	0.000	0.994	1.005
ASSI	-1.438	12.026	0.001	0.237
Road Proximity A road	0.206	10.149	0.001	1.228
Road Proximity B road	0.181	9.477	0.002	1.190
Sewage constraint	0.176	2.078	0.149	1.192
Flood event area	-0.451	3.254	0.071	0.637
AONB	-0.164	6.778	0.009	0.849
Settle Band 1	0.454	23.154	0.000	1.575
Settle Band 2	0.087	0.071	0.789	1.091
Settle Band 3	0.534	3.945	0.047	1.705
Settle Band 4	-0.093	0.566	0.452	0.912
Settle Band 5	0.099	0.395	0.530	1.104
Settle Band 6	0.348	5.705	0.017	1.416
Settle Band 7	-0.021	0.012	0.913	0.979
Settle Band 8	0.055	0.138	0.710	1.057
Constant	-1.895	0.000	1.000	0.150
Model χ^2	3956.42***			
-2 Log likelihood	15257.86			
Cox & Snell R^2	0.233			
Nagelkerke R^2	0.322			
Hosmer and Lemeshow χ^2	14.210			

When incorporating the 30 week determination threshold into the logistic regression framework, the findings show that there is a general reduced effect in terms of magnitude of the odds ratio (Table A5). For example, the number of units coefficient is nominally negative with the odds ratio effectively equating to 1 which indicates no change in the probability of being within the target group (30 weeks) as the score on the predictor changes. In essence, the unitary increases in unit density show no probabilistic likelihood of falling outside the 30 week threshold, which suggests that beyond this period there is no decreasing likelihood that the increase in the number of units comprise longer processing time. When examining the unit density categories, whilst the coefficients remain negative and symbolise that there is a decreasing probability that they will be within the 30 week processing period, this has reduced – as expected, yet does indicate that unit density of applications impact upon processing time, albeit not at any conventional statistical level. Indeed, the findings largely reflect statistical insignificance across the remaining planning policy and infrastructure characteristics. The analysis therefore indicates that beyond the 30 week processing period the effects across the characteristics within the data are randomised.

Table A5: Logistic regression coefficients for 15 week processing

	β	Wald	Sig.	Exp(β) OR
Number of Units	-0.010	3.144	0.076	0.990
Unit density 1-4	-0.237	0.170	0.680	0.789
Unit density 5-24	-0.634	1.385	0.239	0.531
Unit density 25-49	-0.938	3.759	0.053	0.392
Conservation area	-0.202	0.943	0.332	0.817
Landscape policy area	-0.200	2.235	0.135	0.819
Special conservation area [SAC]	-0.035	0.005	0.946	0.965
ASSI	0.029	0.010	0.920	1.030
Road Proximity A road	0.206	10.149	0.001	1.228
Road Proximity B road	0.181	9.477	0.002	1.190
Sewage constraint	0.026	0.061	0.805	1.026
Flood event	0.139	0.426	0.514	1.149
ANOB	0.035	0.369	0.544	1.035
Settle Band 1	-0.079	0.878	0.349	0.924
Settle Band 2	0.297	0.873	0.350	1.346
Settle Band 3	0.185	0.613	0.433	1.203
Settle Band 4	0.112	1.005	0.316	1.118
Settle Band 5	-0.136	0.966	0.326	0.873
Settle Band 6	-0.161	1.609	0.205	0.852
Settle Band 7	-0.141	0.738	0.390	0.869
Settle Band 8	-0.129	0.980	0.322	0.879
Constant	-0.985	0.000	1.000	0.373
Model χ^2	1236.62***			
-2 Log likelihood	16908.62			
Cox & Snell R^2	0.079			
Nagelkerke R^2	0.113			
Hosmer and Lemeshow χ^2	1.4210			

Key findings

The findings emerging from the modelling exercises exhibit evidence which would be generally expected and help confirm, in a statistical sense, the anecdotal evidence surfacing throughout the gestation of this research. As expected, the data shows there to be no 'typical' planning application with ongoing encumbering features serving to impact upon the magnitude and the degree of statistical (in)significance. The regression analysis at the overall NI level suggests that there does not appear to be any real trend/issue regarding time taken from original receipt of a submitted application to validate the application. As highlighted in the descriptive analysis, this occurs for 98.5% of applications within 10 days. Where this does not occur, it is showing it to be a significant predictor of processing time, namely, if an application takes longer than 21 days to validate this inferably increases the time for determination and has obvious inherent features which will delay processing time. Whilst axiomatic, the number of units consistently appears, in its continuous measurement state, as comprising an effect on increasing delays. Nonetheless, this is not omnipresent when examining this aspect both temporally or spatially – indicating that at some points in time this has been more pressing and it is more of a concern in some LGDs than others. Moreover, this appears to have alleviated more recently based on 2017 and 2018 data - and in a statistical manner no longer seemingly a prevalent concern. The categorisation of the unit density also shows a consistent effect on determination time within the standard regression model, however, this effect is not uniform or significant across a number of local government districts and generally beyond the 30 week processing time. What has emerged from the binary categorisations is the 50+ units are not always increasing application time, on the contrary they appear in regional sense not. In 2018 they revealed a large negative effect (on the model intercept) showing them to be creasing the processing time. When considering this for each spatial (LGD) model this effect also holds true. This is also confirmed in the logistic analysis which shows unit density to have a sizeably reduced effect beyond the 30 week processing time periods. Estimation using the probabilistic odds ratio does however show that unit density is primarily a concern for realising the 15 week timeframe and indeed whether an application conforms with conservation and AONB designation appears to of concern. The challenge in a modelling sense – in relation to the detail and granularity of the data – is the randomness of the applications and the uniqueness (and complexity) of them spatially (including the vertical nature [rurality versus urban level]), and temporally is reflected in the low levels of model explanation. In essence, there does appear to be some prevalence in a few characteristics however for meaningful evidence-based policy making the consistent significance and levels of these effects would need to be more acute, profound and insightful.