MARCH 2019

Homelessness

Causes of Homelessness and Rough Sleeping

Feasibility study

alma economics

Homelessness

Causes of Homelessness and Rough Sleeping

Feasibility study

Table of Contents

Non-technical summary

The feasibility study explores a set of available options for a suite of models on homelessness and rough sleeping in England. This is part of a three series project that includes a rapid evidence assessment of homelessness and rough sleeping causes in the UK and abroad as well as a review of existing models on homelessness. Key findings from these strands of research inform drive our recommendations for developing models to estimate future trends in homelessness and rough sleeping and appraise government policies.

Specifically, the models could be used by analysts in the Ministry of Housing, Community and Local Government (MHCLG) and the Department for Work and Pensions (DWP) to address the following objectives:

- generate accurate short-term forecasts of various types of homelessness including statutory homelessness,[1] single people homelessness and rough sleeping
- project medium to long term trends in various types of homelessness
- appraise the impact of suggested policy changes.

Key choices for the development of a suite of models on homelessness

The key message from assessing the characteristics of different classes of models on homelessness is that each model class has specific aspects that render it more suitable for certain purposes than others. Based on this finding, we recommend the development of a suite of different models to address each distinct objective rather than a single, multi-purpose model. The suggested suite should include the following:

- time-series models for accurate short-term forecasts
- simple, ad-hoc simulation models for appraisal of specific policies
- complex simulation models for medium to long term projections of homelessness types conditional on a broad set of predictive factors that are shown in the literature to influence homelessness.

Time series models

The optimal solution for predicting levels of homelessness and rough sleeping in the short-term is the development of time series models that are empirically shown to generate

[1] Under the Homelessness Reduction Act 2017, the definition of statutory homelessness has been recently extended to include all homeless people (including single homeless and those in hidden homelessness) who turn to Local Authorities for homelessness and rough sleeping services. For ease of reference and to avoid any confusion when putting this report into context, we use the term ‘statutory homelessness’ to refer to the former official definition (i.e. homeless households that are accepted as homeless and in priority need by LAs, which is still universally used in the literature on homelessness and rough sleeping in England.

accurate predictions in the near future. The models are simple in that they arrive at short-term forecasts based on historical trends and are not dependent on factors that are shown to predict or cause homelessness and rough sleeping. While there are versions of time series models that include a set of predictive factors and can be used to evaluate the impact of policy changes, they are not an optimal method for policy appraisal as they won't correctly identify the relationships from predicting factors to homelessness.

Simulation models

Economics-based simulation models project outcomes of interest conditional on a set of predictive factors. They are based on a solid theoretical framework that allows for modelling homelessness and rough sleeping as the outcome of complex relationships between a broad set of predicting factors. In theory, the models can produce short-term predictions – however, their outputs depend on estimated future trends and potential relationships between a broad set of determinants that are likely to materialise in the medium to long term. Therefore, these type of models are better placed for appraising policies and estimating long-term trends rather than producing predictions in the short-term.

In the long run, it is important that MHCLG and DWP can use complex simulation models to conduct a comprehensive analysis of the mechanisms driving future homelessness and rough sleeping trends. However, the development of such a model is a long-term process that requires high levels of expertise and substantial investment in resources.

In the short-term, we suggest the development of simple, ad hoc simulation models to provide timely evidence-based assessments of future policies. These simpler versions of economics-based simulation models can be used to help quantify the net effects from introducing new policies without having to consider baseline trends in homelessness (in the absence of the policy) and the factors that drive them.

Data inputs and model elements

The rapid evidence review of the causes of homelessness and rough sleeping revealed that homelessness is a complex phenomenon that emerges as a result of intricate interactions between a broad set of policy, economic and personal factors. Policy analysts can choose the set of predictive factors that should be included in the models based on the model's objectives.

For example, time series models can generate forecasts simply by using historical series of data for the variable of interest. Simple ad hoc models can include a set of variables that are relevant to the policy in question while more complex simulation models usually integrate a number of modules to model the links between outcomes of interest and a broad set of explanatory factors.

The suggested models can be developed using existing sources of data on homelessness and predictive factors – i.e. administrative sources of data on homelessness and rough sleeping collected by LAs, data from surveys that are either centred around homelessness or include relevant information, and official statistics for predictive factors. However, better data can result in more reliable outputs using the same methodology. For example, more granular and precise estimates of different types of homelessness can be achieved if the following data improvements are realised:

- covering different types of homelessness (e.g. sofa-surfing, overcrowding),
- linking data from various sources, and
- improving consistency and data sharing across LAs.

As reliability of outputs depends on quality of available data, improvements in data on homelessness are considered to be of equal priority to development of robust models.

Key modelling choices

Patterns of homelessness and rough sleeping vary from place to place across England and are likely driven by interactions between a range of different factors specific to each area. In this context, it is likely that national policies around homelessness and rough sleeping have different impacts throughout England. Therefore, regional variation is a critical aspect to consider in developing a suite of models on homelessness in England. Moreover, the literature suggests that experiences of homelessness differ across vulnerable groups. Therefore, it is important to understand the impact of national policies on different segments of the population (e.g. low-income households, victims of domestic violence, immigrants, people with mental health and drug abuse problems).

It is important that the suite of models can produce **highly granular outputs** across different levels of geographic disaggregation (e.g. regions, local authorities), types of homelessness that are driven by different underlying factors (e.g. sofa surfers, concealed homelessness) and population groups. The entire set of components of the model suite should be developed using detailed data that allow for disaggregation at the geographical level and across different population segments.

Moreover, the suggested suite of models should **be easy to use and maintain** by in-house analysts. While the development of some elements of the model suite could be externally commissioned (e.g. time series model and complex simulation model), the departments should be able to operate, revise and update the entire set of components of the model suite using their own resources and expertise.

Documentation, including guidance for model applications as well as front ends that will allow users to easily operate the models, should be provided along with the core models. The departments should also invest time and resources to **train in-house analysts to revise and update the models** – for example, using updated data or different assumptions about key model parameters. It should be noted here that developing an easily-accessible front end and detailed guide can often be as difficult and time consuming as developing a model's core 'engine'.

In the case of simulation models, implementing a full modular structure is important to ensure that even a complex model can be accessible to in-house analysts. A large model should build in separate components explicitly considering and planning for future adjustments in the development stage. It is important that the separate modules are built in a consistent way that allows different teams of analysts to revise the model or add new modules without having to change the core model structure.

Finally, the development of a suite of models that produces highly granular outputs and considers the impact of broad sets of determinants is a demanding and long-term process. Therefore, it is important that the departments develop or retain the expertise to design and use ad hoc simulation models that consider a limited set of links between predictive factors and outcomes of interest to conduct ex ante evaluation of potential policy changes within limited time frames.

1. Scope

1.1 Overview and aims

This feasibility study seeks to explore options for the development of a model, or a suite of models, that could be used to assess the impacts of Government intervention on levels of homelessness.

This study is informed by three strands of research that were conducted as a part of the wider feasibility project:[2]

- a rapid evidence assessment of the factors that cause various types of homelessness in the UK and overseas,
- a review and assessment of the suitability of existing methodologies that have been applied to accommodate different policy purposes related to homelessness, and
- an overview of existing data and recommendations regarding potential areas of improvement for data that feeds into homelessness models.

The findings of these exercises provide an evidence base for identifying available options for developing a suite of models to predict homelessness trends in the future and appraise planned changes in broad policy areas. Available evidence regarding data inputs, modelling options and explanatory factors will guide our recommendations about the development of methodologies suitable for addressing distinct national policy questions in the most effective way.

Instead of focusing on a single and complex model that can potentially address all different objectives, we propose the development of a composite model suite that comprises various components. In our recommendations, we take into account an array of issues related to data inputs, outputs, resources, methodological considerations, modelling choices and types of policies that the models should consider.

This section outlines the purposes that the Ministry of Housing, Communities and Local Government (MHCLG) and the Department for Work and Pensions (DWP) seek to accommodate by conducting empirical research. Moreover, we identify and discuss a range of options for adapting approaches to address these objectives. Section 2 maps existing sources of administrative and survey data on homelessness as well other types of data that can potentially

[2] Findings from the rapid evidence assessment on causes of homelessness are presented in a separate report under the title “Rapid Evidence Assessment”. The review and assessment of existing classes of models used to predict and measure homelessness is presented in a separate report under the title “Review of models of homelessness”. The overview of existing data, evidence gaps and recommendations for potential areas for improvements in collections of data on homelessness is presented in section 2 of the current report.

feed into homelessness models, summarises upcoming data collections and highlights gaps in the evidence to make suggestions about potential improvements in data collection. Section 3 discusses a set of key modelling choices that apply to the entire collection of models that should be developed while section 4 highlights modelling issues that are specific to the particular components of the model suite. Finally, section 5 sets out recommendations and considerations for developing a suite of models to inform policies on homelessness.

1.2 Purposes of the model suite

The model suite is meant to be used by MHCLG and DWP to address the following set of objectives related to various types of homelessness and rough sleeping:

- short-term forecasting,
- projections of medium to longer-term trends, and
- appraisal of hypothetical policy scenarios designed to influence levels of homelessness.

The discussion is centred around the development of an empirical approach to forecasting and projecting future levels of different types of homelessness types under baseline assumptions or alternative policy scenarios that are likely to affect homelessness (for example, policies affecting the supply of housing or levels of welfare support for housing costs).

We will discuss the applicability of different classes of models in addressing these three distinct objectives for various types of homelessness. Theoretically, a methodological approach applies equally well to all outcomes related to homelessness since these outcomes are measured by variables of the same type (e.g. continuous variables for population counts, probabilities for estimation of homelessness risks, etc.) For example, the same time series model (e.g. Autoreggressive Integrated Moving Average – ARIMA – model) can handle different series of data inputs to forecast the entire range of homelessness types (for example, inclusing single homeless, and sofa surfers) and rough sleeping. What might change across different types of homelessness are the assumptions about model determinants in the sense that each homelessness type is likely to be driven by different mixtures of causal and predicting factors.

Our objective is not just to recommend ways for producing forecasts and projections for broad categories of homelessness such as statutory homelessness, single homeless and people who sleep rough. Instead, we will identify a range of options for generating outputs at different levels of disaggregation (e.g. new homelessness levels among former care leavers, homelessness among black and minority ethnic (BME) groups, returns to rough sleeping among people with complex needs).

Box 1. Types of homelessness – a discussion about definitions

According to the Office for National Statistics (ONS) and MHCLG, a household or an individual is considered homeless and can apply for homelessness support when they:

"*no longer have a legal right to occupy their accommodation or if it would no longer be reasonable to continue to live there, for example if living there would lead to violence against them".*

Moreover, the official MHCLG definition for people who sleep rough is the following:

People sleeping, about to bed down (sitting on/in or standing next to their bedding) or actually bedded down in the open air (such as on the streets, in tents, doorways, parks, bus shelters or encampments). People in buildings or other places not designed for habitation (such as stairwells, barns, sheds, car parks, cars, derelict boats, stations, or "bashes").

The rough sleeping definition does not include people in hostels or shelters, people in campsites or other sites used for recreational purposes or organised protest, squatters or travellers.

Prior to the recent Homelessness Reduction Act 2017, the official definition of statutory homelessness comprised three criteria:

- being eligible for assistance,
- being unintentionally homeless – 'intentionally homeless' are considered the households that left a home that could have stayed in, and
- falling within a specified priority need group – households with dependent children or a pregnant woman; individuals who are vulnerable as a result of mental illness or physical disbaility ; individuals aged 16-17 years old; individuals aged 17-19 who were previously in care; vulnerable individuals as a result of previously being in care, HM forces or under custody; vulnerable individuals who had to flee their home as a result of violence or threat of violence.

The Housing Act 1996 –provides that where an applicant meets the above three criteria, then local authorities (LAs) have a statutory duty to provide them with a settled home, and where this is not possible straight away, they are under a duty to provide suitable temporary accommodation until settled accommodation can be found. Counts of households and people that were in temporary accommodation following accepted homelessness applications were reported at the end of each quarter. National statistics on statutory homelessness were derived from these counts reported by LAs. While rejected applications for homelessness (either because households were found not to be in priority need or because they were considered to be intentionally homeless) were also reported, no other information on these groups that are considered to be non-statutory homeless was reported.

According to MHCLG, there are three sub-groups in the non-statutory homelessness category:

- single homeless,
- people who sleep rough – people bedded down in the open air, and
- hidden homeless – people who are homelessness but are not visible in official statistics (sofa-surfing).

The new Homelessness Reduction Act 2017, which came into force in April 2018, leads to important changes in the delivery of homelessness services. Under the new Act, LAs are required to offer two new duties (prevention and relief) to all applicants that are eligible even if they are intentionally homeless or do not fall into any priority needs category.

In this context, the new official definition for statutory homelessness has been broadened to include the entire range of single people and households that apply to the LAs for homelessness support (even if they are not eligible for temporary accommodation). Therefore, the new national statistics need to integrate figures on what previously was considered non-statutory homelessness in addition to rough sleeping.

Developing a broader definition is critical for guiding collection of data that cover the entire range of homelessness types including statutory homelessness, rough sleeping, sofa surfing and concealed homelessness ('over-crowding').

Bramley (2017) suggests the following two alternative definitions that are broader in the sense that they integrate forms of non-statutory homelessness that fall out of official statistics:

- Core homelessness that includes the most acute forms of homelessness (rough sleeping, sleeping in tents and cars, unlicensed and insecure squatting, unsuitable, non-residential accommodation, hostel residents, users of night/winter shelters, domestic violence victims in refuge, unsuitable temporary accommodation, sofa surfing), and
- Wider homelessness that refers to people who are at risk of homelessness or stay in some form of temporary accommodation (staying with friends and relatives due to inability to find proper accommodation, eviction/under notice to quit, asked to leave by parents/relatives, intermediate accommodation and receiving support, in other temporary accommodation, discharged from prison, hospital or other state institution without permanent housing).

Finally, efforts have been made to establish a harmonised official definition of homelessness across the UK. The Government Statistical Service (GSS) Harmonisation Team, which is part of ONS, has been recently commissioned by MHCLG to map the definitions of homelessness that are used in the UK and investigate options for developing a harmonised homelessness definition.

It was found that different homelessness definitions reflect differences in homelessness policies and priorities in delivery of prevention and support services across the UK countries. Moreover, information regarding the comparability between different definitions appears to be limited.

The GSS harmonisation team has further explored a set of homelessness definitions that are used across government bodies (e.g. MHCLG for national statistics on homelessness, DWP for those in need of benefits and the Ministry of Justice (MoJ) for assessing accommodation of ex-offenders) and non-government organisations (for example, core and wider homelessness definitions used for CRISIS projections of future trends in homelessness).

Variations in homelessness legislation and operational differences when applying the definitions to produce homelessness statistics were also examined across UK countries.

Findings from this research revealed that introducing a harmonised definition would require changes in legislation and data collections across the devolved nations that are not straightforward to implement. Therefore, the GSS harmonisation team recommended that a conceptual framework for homelessness should be created in order to map different definitions and data collections in the UK and improve comparability of existing statistics.

1.3 Modelling options

Accounting for the complexity of the phenomenon under analysis and the theory underpinning homelessness as well as distinguishing between the distinct purposes of short-term forecasting, long-term projections and policy appraisal are key considerations when developing models around homelessness.

For instance, the complexity of relationships between different factors – such as the interconnections between the housing and labour markets across English areas – that influence homelessness levels is an important element that should be considered when projecting long-term homelessness trends. However, including assumptions about such relationships to estimate trends in the short-term is likely to result in decreased forecasting accuracy.

A key choice that needs to be made is between developing a single, large-scale and complex model that integrates multiple features or a suite of simpler models that are used to accommodate distinct purposes. A number of issues, including the applicability of the available methodologies as well as the costs associated with each option, should be considered in informing the choice of the optimal strategy.

Specifically, the costs of developing and using a large-scale, complex model that integrates various features to model all possible links and interdependencies between related factors and homelessness types might exceed the benefits of having a single model that can address all purposes. Moreover, the development of a complex comprehensive model requires time and resources while smaller models can be designed in the short term to address immediate policy objectives. The design of small ad hoc models can also be seen as a critical step to the long-run process of developing a robust complex simulation model that can be used for the entire set of policy purposes associated with homelessness and rough sleeping.

For example, the CLG-Affordability model – a complex simulation model that estimates housing affordability in England as the outcome of a number of interconnected determinants (NHPAU, 2009) – consists of a set of simpler simulation models on house prices as well as housing demand and supply that can be used separately. These models have been also utilised in the development of the components of the Sub-Regional Housing Market Model (SRHMM) developed by Bramley and Watkins (2016).

In the review and assessment of classes of models that are used to measure and predict homelessness, we identified a set of methodologies that can be applied to address policy purposes around homelessness. The key take-away from the model review and assessment is that there is merit to applying different models for different purposes.

As shown in figure 1, which summarises the main findings of the model review and assessment, each class has particular statistical properties that makes it more suitable for some purposes than for others. For example, time-series models are simple trend-based methods that generate accurate forecasts of outcomes of interest in the short-term based on the underlying assumption that patterns that existed in the past will continue into the future. While they can be applied to estimate medium to longer-term trends, they lack the theoretical framework that is needed to account for relationships between explanatory factors and outcomes that play out in the long-term.

Based on our findings[3] and the above discussion, we recommend the development of a flexible suite of models that will comprise a set of methodologies applied to address different objectives instead of a complex, large-scale model. Specifically, we suggest that models from the following two broad classes should be applied to accommodate MHCLG and DWP policy objectives:

- time-series models for short-term forecasting,[4] and
- economics-based simulation models for medium to long-term projections and policy appraisal.

Projections of trends under baseline assumptions and evaluation of changes in homelessness levels under alternative policy scenarios are separate exercises.

Though these objectives can be covered within the same overall framework (for example, a simulation model can do both), different versions of models that fall within this class can be used to address these two distinct objectives.

Essentially, simple ad hoc models can be used to quantify the impact of specific policies compared to a baseline *'do nothing'* scenario. The estimation of additional effects from launching a new policy does not necessarily require considering the baseline levels of homelessness and rough sleeping. SRHMM (Bramley and Watkins, 2016) is an example of a comprehensive simulation model that projects housing needs, including homelessness, under composite policy and economic scenarios.

[3] Our recommendations for suitable models are based on findings from reviewing and assessing the set of model classes that have been used to predict and measure types of homelessness. For a detailed discussion about the characteristics of existing models, see the "Review of model of homelessness" report.

[4] As discussed in the model review and shown in figure 1, machine learning techniques are an alternative option for generating accurate short-term forecasts. However, the reliability of machine learning outputs relies on the amount and level of detail of data on homelessness. Therefore, applying such models to English data (facing a number of limitations that are discussed in section 2 of this report) is likely to be suboptimal.

Box 2. Other policy objectives and methods to address them

The following policy purposes can be also addressed by applying empirical models:

- identifying homelessness risks for households and individuals and single people,
- measuring homeless groups that are not straightforward to capture, and
- evaluating existing policy interventions that address homelessness.

The models used to accommodate these purposes include:

- homelessness risk models,
- non-standard sampling models such as the capture-recapture method, and
- models developed to quantify intervention (treatment) effects for participants in the period following the intervention.

The focus of this feasibility study is not to recommend ways to explicitly address these additional objectives. However, the above methods can potentially complement the main models developed to predict homelessness levels and appraise policies. For example, outputs from homelessness risk models can be integrated into larger and more complex policy models that simulate homelessness outcomes under different scenarios.

Alternatively, these methods can be used as stand-alone policy tools developed outside the main models. It may be worthwhile for MHCLG to sponsor a project that brings together expertise from LAs that use homelessness risk models to develop a common approach to identifying households and single people that are in priority need for homelessness prevention services. Such an approach can also be adopted to assess differential impacts of the implementation of central policies in different UK regions. For example, policies around private rent prices (e.g. housing benefits) and housing supply (e.g. investment in council housing) are expected to exert significant impact on homelessness risks in London Boroughs where high private rents and shortages in supply of council housing are important drivers of homelessness. On the other hand, such policies are not expected to have a similar impact in Northern England where access to social housing does not appear to be a major issue (Fitzpatrick et al. 2018).

Capture-recapture methods, which use a set of sampling techniques to estimate the size of populations that are elusive and thus not straightforward to measure,[1] can be implemented to guide new data collection that can improve the outputs of models. Alternatively, these methods can be applied to existing data to produce reliable counts of populations that are not easy to measure such as sofa surfers and households in concealed homelessness, improving area-based counts of homelessness groups.

Finally, ad hoc models that identify treatment effects can be used to assess the effectiveness of previous or existing interventions. For example, a similar strategy was adopted to measure the impacts of realised changes in Local Housing Allowance (LHA) on a number of outcomes, including LHA entitlements, contractual rents and types of properties claimants live in. (Beatty et al., 2014). A difference-in-differences model was applied to administrative data on housing benefits claims from the Single Housing Benefit Extract (SHBE) to compare trends in outcomes (for example, rents and types of properties) for groups who moved into the new LHA system to groups with similar characteristics that have not rolled onto the new system yet.[2]

Notes

[1] For a more detailed discussion about capture-recapture methods see the “Review of models of homelessness” report.

[2] For a comprehensive outline of the model developed to measure the impact of LHA reforms, see the report by Brewer et al. (2014).

2. Collections of data

2.1 Overview

There are three types of data which can be used to project homelessness and rough sleeping in the future and evaluate the effects of policies aimed at tackling homelessness and supporting people in need: [5]

- administrative data on homelessness and rough sleeping collected by LAs and reported by MHCLG at frequent time intervals,
- data at the household and/or individual level from large scale household surveys which include information on homelessness or surveys that were designed to explicitly cover homelessness and rough sleeping experiences, and
- administrative data (for example, official statistics) on homelessness determinants – e.g. housing and unemployment benefits, housing supply, private rents, demographic trends, health indicators, key economic variables, etc.

Time series models can be applied to series of administrative data on homelessness and rough sleeping that are reported frequently (e.g. every quarter). These models can handle large series of data inputs (for example, across LAs and for particular population groups) to produce granular short-term forecasts.

Previous collections of administrative data on homelessness (collected using P1E forms for people in temporary accommodation), which are aggregated at the local authority level, included a limited set of background information. The Homelessness Case Level Information Classification (H-CLIC) system for data collection which will replaces the P1E forms, collects household case level data providing more detailed information on the causes and impacts of homelessness, long-term outcomes for homeless households and what works best for preventing homelessness. Moreover, administrative data on rough sleeping collected using the Rough Sleeping Evaluation Questionnaire (RSEQ) include information on individual socio-economic characteristics that have been shown to be associated with homelessness (e.g. financial strain, use of other public services, mental health problems, etc.) Therefore, forthcoming collections of administrative data can be used to estimate time series models that include limited sets of explanatory variables in addition to historical values of the variables of interest (multivariate models).

H-CLIC and RSEQ data can be also used to measure the effects of predictive factors on different types of homelessness at the first stage of simulation models. Survey data and other sources of statistics on key

[5] A detailed overview of existing data sources can be found in the appendix.

determinants can also useful for estimating homelessness projections using simulation models. Individual data drawn from surveys can feed into components of simulation models to quantify behavioural responses to changes in important predictive factors. Detailed individual data from surveys are necessary for developing models that can produce granular outputs – e.g. micro-simulation components that produce distributional outcomes or generate projections for subsamples with specific characteristics. Data on other explanatory variables are also used in simulation models to arrive at homelessness projections conditional on future trends in determinants.[6]

In principle, selecting suitable methodologies to project outcomes of interest and evaluate policies does not depend on data availability and quality in the sense that there is no merit in developing different methodologies for different data. Analysts select the methods they will use from a set of existing options and rely on available data to accommodate policy objectives. When data is imperfect or not available, they make assumptions to address the limitations imposed by lack of data or data of low quality.

For example, when detailed data on other life domains of homeless people are not available, assumptions are used to compensate for missing knowledge about personal characteristics that might influence paths in and out of homelessness. Such assumptions (including the extrapolation of missing predictors using other observable characteristics) might lead to decreased output accuracy. For instance, if data is not available on a predictor that is highly correlated with homelessness such as income, we would have to use an observable proxy such as socio-economic status or educational achievement to approximate individual income. We would then quantify the link from income to homelessness based on this approximation, which would result in reliability losses in our homelessness estimates conditional on income.

In the case of homelessness and rough sleeping in England, existing sources of data on outcomes of interest and their determinants are adequate for applying models to predict future levels of homelessness under composite policy and economic scenarios. However, improving the quality of existing data or collecting new detailed data on homelessness and rough sleeping will certainly influence the model outputs – more detailed data lead to more reliable outputs under the same empirical design.

2.2 Evidence gaps and areas for improvement

In this section, we highlight potential areas for improvement in data collection based on gaps that we have

[6] See box A1 in the appendix for forecasts of predictive factors.

identified in existing data sources on homelessness and rough sleeping in England. New collections of data and enhancements to already existing systems for gathering information will result in more reliable estimates of future trends in homelessness under alternative policy and economic scenarios. More detailed evidence of homelessness experiences at the individual level will contribute to a better understanding of the causes and impacts of homelessness as well as what works best for preventing and reducing homelessness.

We also consider the importance of suggested enhancements in data as part of developing a comprehensive evidence base that will result in more reliable estimates of homelessness and rough sleeping. We categorise them in two groups:

- top priority – data that are necessary for generating robust projections of various homelessness types, and
- further priority – data that can add depth but are not central to achieving the aims of a suite of models around homelessness.

2.2.1 Top priority

In this section, we discuss recommendations for improving data on homelessness that are critical to conducting a robust empirical analysis of homelessness trends, pathways in and out of homelessness and the contribution of broad policy areas to reduction and preventions of homelessness.

Covering all homelessness types

The new Homelessness Reduction Act 2017 required local authorities (LA) to meet two new duties (relief and prevention) to all those affected, regardless of priority need or intentionality.[7]

Following this major change in policy, it is important that LAs gather information about types of homelessness in addition to the former definition of statutory homelessness – for example, sofa surfing, squatting and living in hostels and other types of short-term or emergency accommodation.

The development of comprehensive definitions of various homelessness types is central to the design of a systematic recording of homelessness types that covers the entire range of homelessness experiences in England – for instance, single people homelessness, rough sleeping and sofa surfing. A common and comprehensive description of what homelessness is and which groups of people are owed support by public services in England will guide the collection of consistent data on homelessness outcomes of interest.

Examples of collecting data on various homelessness types are the additional modules and questions included in the Rough Sleeping Questionnaire as well

[7] For a detailed discussion about the Act see here: http://www.legislation.gov.uk/ukpga/2017/13/contents/enacted

as well as in the H-CLIC form for data collection. The Rough Sleeping Questionnaire includes questions that capture past experiences of sofa surfing in addition to rough sleeping. It could be administered to all local authorities in England and become part of official statistics. Moreover, collecting data at regular time intervals – for example, in annual or bi-annual waves – as well as adding a longitudinal element to data collection would improve statistics on rough sleeping and contribute to a better understanding of individual experiences. The H-CLIC form is completed by all local authorities and includes a question about last settled accommodation and type of accommodation at the time of the application.[8]

Data linking

Using datasets that comprise linked administrative data from distinct sources that cover large numbers of areas (e.g. benefits, health, institutional history) is an important tool for research that aims to understand complex social issues and inform policy. It allows for capturing links between a broad set of predictors and outcomes of interest and mapping the array of paths to the incidence of social problems such as homelessness.

Poor linkage of data in the English context is a major limitation to a comprehensive analysis of homelessness that could contribute to a better understanding of the problem, its causes at the personal, economic and policy level and what policies are needed to tackle it.

Administrative data covering a number of areas including welfare benefits, health and use of public services can be linked to other administrative data on homelessness and rough sleeping. For example, the Single Housing Benefit Extract (SHBE) dataset, collected from LA records, is the key administrative source of monthly data on housing benefits claim. This contains data on household type and demographic characteristics, amount of monthly rent, share of the rent that is covered by Local Housing Allowance and type of accommodation. Linking such benefit data to data on people who are either homeless or at risk of homelessness would allow analysts to identify the contribution of housing benefits to homelessness prevention.

[8] LA homelessness services applicants are asked about the type of their accommodation at the time of the application. They can choose between: owner-occupier; shared ownership; private rented sector; council tenant; registered provider tenant; Armed Forces accommodation; tied accommodation, looked after children replacement; living with family; living with friends; social rented supported housing (or hostel); refuge; rough sleeping; homeless on departure from institution (custody/hospital); temporary accommodation; student accommodation; National Asylum Support Service accommodation; no fixed adobe; caravan/houseboat. In the cases where the applicants report that their current accommodation is not their last settled home, they are asked about their accommodation when they were last settled in order to capture routes into homelessness. The applicants can choose between owner/occupier; shared ownership; private rented sector; lodging (not with family/friends); council tenant; registered Provider tenant; living with family or friends; looked after children placement; social rented or supported housing; tied accommodation; Armed Forces accommodation.

For example, the Public Health Outcomes Framework sets out desirable health outcomes at the national and subnational level and measures health indicators across LAs in England. The dataset also includes two indicators on homelessness that potentially allow for modelling links between physical and mental health outcomes and homelessness at the LA level. However, the indicators only capture statutory homelessness at the LA level, hindering the assessment of links between health outcomes and other types of homelessness – such as sofa surfing and rough sleeping. Given that mental health appears to be a major determinant of rough sleeping, there is merit in expanding the accommodation response category in the Public Health Outcomes Framework to capture other types of homelessness and link the observations to official statistics on rough sleeping or administrative data collected using the Rough Sleeping Questionnaire.

Other sources of data that contain information on accommodation types, including homelessness, that could be linked to homelessness data, such as H-CLIC, are the following:

- data on care leavers aged 17-21 years old drawn by Children Looked After in England,
- data on prisoners drawn by Accommodation Status of Prisoners and Police Records,
- data on groups of drug treatment services drawn by National Drug Treatment Monitoring System, and
- data on groups who are vulnerable because of physical and mental health issues drawn by Mortality Statistics, Mental Health Minimum dataset, Hospital Episode Statistics and the Health Improvement Network.

There are various considerations concerning issues related to technical and legal aspects of the data linking process. An important issue is anonymisation of data and security of information. Explicit guidelines and protocols should be put in place to ensure that it is not possible for analysts who use the dataset to link data to people. For example, the number of attributes included in the compilation of administrative data is an issue to consider – a wide variety of attributes could lead to the identification of specific service users in small LAs, where limited numbers of people experience homelessness. Despite the variety of issues that need to be considered, linking existing sources of data could be a more straightforward and less costly – in both resources and time – alternative to expanding existing data sources or designing new collections to capture additional information about people who are either homeless or at high risk of homelessness.

Box 3. Steps toward data linking in England: the Rough Sleeping Evaluation Questionnaire (RSEQ)

The Rough Sleeping Evaluation Questionnaire (RSEQ) was introduced as part of the recent MHCLG initiative to tackle the most severe form of homelessness – i.e. rough sleeping. The new instrument for data collection contributes to existing approaches by collecting detailed data on individuals' past and current experiences of rough sleeping and capturing a wider set of factors that are related to such experiences, including support needs, feelings and attitudes and health indicators.

In addition to this contribution, the new method goes beyond prior approaches to data collection by proposing a scheme for data linking across administrative datasets. Personal details of service users interviewed with the RSEQ – such as names, date of birth, and national insurance number (if known) – are linked to:

- administrative data on receipt of welfare benefits (DWP),
- criminal justice system records (MoJ),
- administrative data on statutory homelessness applications collected by LAs (MHCLG),
- health care services use (NHS Digital),[1] and
- alcohol and drug treatment use (PHE).

The output of this process is a comprehensive dataset that includes detailed information about a broad set of areas – history of rough sleeping, statutory homelessness applications, support needs, contact with the criminal justice system, receipt of welfare benefits, healthcare use and participation in substance use treatment – but excludes service users' personal details.

Assembling such detailed lists of administrative data for users of homelessness prevention and treatment services is important for understanding the needs of people who sleep rough or are homeless and assessing wider costs of homelessness that potentially exceed the costs of delivery of homelessness services alone.

Notes

[1] *Linking RSEQ data to information on public health service use is likely to be challenging. Evidence from the Homelessness Link survey on the health outcomes of homeless people shows that while 90% of the 2,500 surveyed homeless and rough sleeping individuals are registered with a GP, the rough sleeping groups use GP services the least. See here for more information:* https://www.homeless.org.uk/sites/default/files/site-attachments/The%20unhealthy%20state%20of%20homelessness%20FINAL.pdf

Data from people in households that have been assessed as homeless by Scottish local authorities (LAs) were linked to a number of health datasets covering the following areas: accident and emergency attendance; alcohol-related admissions; drug misuse-related admissions; emergency admissions related to injury and poisoning; psychiatric admissions; and non-attendance at outpatient appointments.

LAs in Scotland do not normally reveal personal information of applicants when sharing data with government departments. For the purpose of this project, all LAs were asked to submit personal identifiable information for people who were considered homeless or at risk of homelessness to the National Records of Scotland (NRS) Indexing Service.

A dataset including personal information about the applicants – such as homelessness application number, name, gender, date of birth, postcode and local authority code – was created particularly for the purpose of data linking. Using the application numbers, this new dataset could be linked back to the homelessness datasets assembled by Scottish LAs.

In order to match homelessness with health data, a 'separation of function' approach was adopted to ensure that no single organisation or individual had access to the entire range of datasets required for this project. A third party (the NRS Indexing Service) matched the homelessness dataset that was created for the purpose of this project with the Research Indexing Spine (RIS) – a population compiled by NRS that uses information drawn from general practitioner (GP) registries at a single point in time (snapshot). The NRS Indexing Service performed the matching only using personal identifiers across the datasets – access to the rest of the data was restricted. Each matched individual was then assigned the Community Health Index (CHI) number that tracks individual usage of health care services.

When matching was completed, the matched results were combined with the rest of the data and the personal identifiers were removed. Analysts accessed this secondary dataset in a separate and secure environment.

Box 4. Best practices in data linkage: the Scottish example

The Scottish government has recently adopted a strategy to promote better use of existing administrative data to understand important social and economic issues and evaluate policies.

Data linking is central to this new approach, which draws on a thorough Data Linkage Framework established in 2012 to promote collaboration and best-practice sharing among key public sector organisations that collect and handle registry data.

A set of guiding principles has been developed to "support the legal, ethical and efficient use of data for linkage purposes within a controlled and secure environment".[1] The principles set out important priorities and considerations related to acting in the public interest, transparency, privacy (consent, anonymisation and security of individual data), data access and consequences when these principles are disregarded.

Efforts have been made to ensure that linked administrative data are anonymised and secure, personal information is protected, and individuals cannot be identified in the datasets. Several anonymisation methods are applied, including complete anonymisation, which excludes all identifiers of personal information from the datasets, and pseudonymisation, where identifying fields (such as names) are replaced with artificial identifiers (such as unique serial numbers). Moreover, safe havens were launched as a way to ensure privacy – these are secure environments where researchers have access only to the anonymised segments of secondary datasets relevant to their research.

Homelessness data linking

One example relevant to analysing homelessness is linking data on homelessness to national-level health datasets. Homelessness data were linked with individual health indicators in order to quantify the use of health services by homeless groups in Scotland (Waugh et al., 2018).

Notes

[1] *For more information about guiding principles for data linkage in Scotland see here: https://www.gov.scot/Topics/Statistics/datalinkageframework/GuidingPrinciples*

Consistency and data sharing across LAs

It is important that LAs collect series of data on homelessness and rough sleeping that go beyond simple counts of groups without fixed abode in a consistent way. The H-CLIC data collection system is a significant step towards the development of a common framework for gathering information on service users that cover a set of topics including demographic characteristics, previous accommodation, other support needs (such as drug abuse and mental health problems) and receipt of benefits.

This new approach should be expanded to the collection of information about other types of homeless groups such as rough sleepers and people in hostels or refuges. The RSEQ that is currently assigned to LAs that offer homelessness prevention services funded by the Social Impact Bond (SIB) and the Rough Sleeping Grant (RSG) programme is a useful tool for surveying service users to collect information on topics such as welfare benefits take-up, housing conditions, income and institutional history. The questionnaire can potentially serve as the basis for the development of a common framework for recording information about broader areas of service users' lives that are related to homelessness experiences.

An efficient system of collection and collation of data depends on good communication and collaboration between LAs. For instance, if working together, LAs will reduce the probability that specific groups of homeless people – such as people who sleep rough – are recorded twice and enhance data accuracy by improving the process of local connections – that is, referral of applicants of homelessness services to other LAs.

Further, data collection practices should be monitored by independent bodies on a regular basis to ensure that LAs follow the same practices and data is consistent and comparable across areas. The guidance and verification provided to LAs by the Homeless Link for counting rough sleeping groups is a representative example of such a process. Independent organisations – for example, homelessness charities such as Crisis – could contribute to monitoring and evaluating the quality, consistency and accuracy of data collected by LAs.

2.2.2 Further priority

In addition to enhancements in data collection necessary for conducting robust empirical analysis of future trends in homelessness and the effectiveness of planned policies aiming to tackle homelessness, we discuss a set of recommendations to improve data on homelessness that would further improve knowledge about homelessness.

Improvements in already existing administrative data

Data that are collected directly from homeless groups who turn to LAs for assistance and advice would benefit from a series of improvements. Below is an indicative list of potential improvements in data collected by LAs:

- more detailed coverage of a set of factors that are shown in the literature to be important predictors of homelessness in the English context (e.g. poverty, domestic violence, relationships with friends and family, use of social services, mental health problems and substance abuse),
- development of a questionnaire that captures policy-relevant predictors of homelessness. For example, when asking homeless people the reasons why they are homeless, the most common response is "due to eviction" (CHAIN, 2018). While this finding is indicative of the factors that lead to homelessness, it is more important to understand the reasons that led to eviction and thus homelessness, that are related to policy - for example, changes to income including benefits, financial strain, shorthold tenancy, etc. The questions included in the new H-CLIC questionnaire on the reasons for loss of settled home, assured shorthold tenancy, social rented tenancy or supported housing (including increases in rent, reduced income from employment, changes in benefit entitlement, etc.) are examples of instruments for collecting data on homelessness determinants that can be affected by policy.
- use of data collection tools – such as the RSEQ – on a larger scale while surveying homeless people (and particularly the most disadvantaged amongst them such as rough sleeping groups). While this is not an easy task, the quality of empirical evidence depends on being able to gather information from large and representative samples, if not the entire population of homeless groups.
- inclusion of direct questions about experiences of homelessness and rough sleeping in large-scale household surveys following representative samples of the population in the UK that cover an array of other life domains (e.g. income, employment, housing conditions, etc.), such as the UKHLS, and
- recording of stocks, flows and returns to types of homelessness instead of single snapshots. While this could be done by adding a longitudinal element to administrative data collection tools (such as the RSEQ and the H-CLIC), it is likely to be challenging as it will pose significant burdens on LAs. Adding retrospective questions to already existing questionnaires to capture past experiences of homelessness and rough sleeping is a more straightforward alternative that could be adopted to measure homelessness flows and returns. There is scope in considering the inclusion of a module for past homelessness and service use in the local data collection pilots that will be introduced in the near future (around summer 2019) as part of

the Rough Sleeping Strategy.[9]

Table 1. *Recommendations for improving collections of data on homelessness*

Recommendation	**Description**	**Issues to consider**	**Priority**
Covering all homelessness types	Use question in H-CLIC data to break down official homelessness statistics by type (e.g. hostel residents, sofa surfers, etc.)	need for harmonised definitions of homelessness	Top
Linkage of data between government sources	Assign a unique identifier to each case to allow for linking data from various sources (e.g. homelessness service users)	data security, anonymisation	Top[2]
Consistency of data collections across LAs	Design of common framework across LAs for data collection and collaboration between LAs	co-ordination between LAs	Top
Improvements in already existing administrative data	More questions regarding the socio-economic background of people who are homeless or at risk (e.g. health indicators, institutional history, use of public services), Information about stocks and flows (and potentially) returns to homelessness Regular updates instead of simple snapshots	costs of expanding current collections potential burden on respondents	Further[2]
Longitudinal data on service use and costs	Common framework for reporting use of services and costs across LAs (additional module to collect data from other public services)	Additional module to already existing framework for homelessness data collection across LAs	Further
Cover hidden populations in surveys	Revise sampling techniques in existing large-scale surveys, design new surveys (or develop extra modules in existing large-scale surveys) to capture hidden populations	Costs (time, effort, expenses) and complexity of developing new approach to survey data collection	Further

[1] Top priority: critical for applying the models, further priority: good to have but not central to empirical analysis.
[2] Linking already existing data from administrative sources can be thought of as a less costly alternative to new data collections, enhancing existing datasets with gathering new information.

[9] See here for more information on MHCLG's Rough Sleeping Strategy: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733421/Rough-Sleeping-Strategy_WEB.pdf

Longitudinal data on homelessness service use and costs

The costs of homelessness services can be found in unit cost databases such as the New Economy Unit (NEU) database.[10] These databases report the average values of financial costs associated with different homelessness services – including temporary accommodation, cost of application, etc. – that are common across local authorities (LAs). It should be noted that these databases are quite out of date – for example, the NEU database relies on temporary accommodation costs reported in 2010/11.

Moreover, while LAs record the costs of the different services they provide to homeless households and single people (e.g. temporary accommodation in hostels and refuges), there is not a systematic way of recording such information.

To conduct comprehensive Cost-Benefit Analysis (CBA) of services offered to populations that are homeless or at imminent risk of homelessness, assess the services' Value for Money (VFM) and estimate total costs of homelessness, data on service use and costs should be collected by English LAs. Small changes in current accounting data series reported to MHCLG by LAs (for example, breaking down housing services costs by type of service – e.g. temporary accommodation) should be easy to implement and contribute to the development of a database detailing homelessness services costs at the LA level.

Moreover, there is scope in introducing modules for reporting service costs as part of local data pilots, which will be launched as part of the Rough Sleeping Strategy to develop and test a multi-agency outcomes framework. For the purpose of consistency and comparability of data, a similar approach could be adopted for homelessness services offered by LAs. The outcome of this approach would be a longitudinal dataset of service uses and costs by type of service that captures LA-specific needs while allowing for assessment of homelessness costs at the national level. It should be noted that the development of such a framework for data collection as well as its implementation across LAs may be a long-term and challenging process, imposing non-negligible burdens on LAs. However, MHCLG can play a central role in coordinating this process to ensure a smooth implementation and minimise financial and other resources costs for LAs (for example, by extending already existing data collections).

While collecting and reporting data on costs at the LA level is an important priority, there are additional steps that could be taken to accommodate the longer-term objective of developing a source of data that allows for a comprehensive assessment of total costs associated with homelessness.

---

[10] http://www.neweconomymanchester.com/our-work/research-evaluation-cost-benefit-analysis/cost-benefit-analysis/unit-cost-database

For example, an additional module could be added to data collection on costs of other public services (for instance, mental health services and drug treatment programmes) in order to identify users who are homeless. Consistent and complete collection of information regarding the wider set of services that homeless groups are likely to use will contribute to better understanding the overall costs associated with homelessness.

Covering hidden populations in large-scale household datasets

Even achieving high levels of detail and quality in the collection of administrative data does not ensure that statistics aiming to count disadvantaged groups and reflect their experiences will be accurate. An intrinsic limitation of administrative data is that they are likely not to accurately capture the true size of the population of interest. For example, people who are in a transient homelessness situation or homeless people who suffer from mental illness might never turn to LAs to seek help and support. On the other hand, there is the possibility that false homelessness duty claims can lead to decreased precision in estimates of homelessness populations.

One possible way to overcome this limitation is the use of survey data to estimate homelessness trends that can be compared against statistics derived using administrative data. Essentially, people are more likely to reveal information about their actual situation when interviewed for a survey where anonymisation of information is ensured. However, surveying disadvantaged groups – such as homeless people – also faces limitations. Homeless populations are hard to contact and survey for a number of reasons – including difficulties to locate them and willingness to participate in surveys.

Overall, UK large-scale household surveys are useful tools for collection of data used to analyse trends and relationships of variables of interest in a variety of connected areas (e.g. health, employment, poverty). While these surveys are designed in such a way to ensure their samples are representative of the populations they aim to analyse, coming up with a representative sample of disadvantaged groups is not always an easy task.

In particular, population subgroups that are at risk of homelessness or have homelessness experiences – currently or in the past – are hard to contact and survey. For example, while the English Housing Survey includes retrospective questions on statutory homelessness, it does not cover past experiences of other types of single homelessness (e.g. sofa surfing). Even if it is straightforward to locate such populations, they may be reluctant to share sensitive personal information during an interview. These limitations are likely to lead to evidence gaps regarding individual experiences of homelessness and rough sleeping and result in decreased reliability of outputs when used as inputs for empirical research aiming to understand and explain homelessness.

There are ways to design sampling in order to best capture vulnerable and disadvantaged groups such as the homeless. For example, Hough et al. (1996) review a set of strategies for how to approach and survey a particularly elusive homeless subgroup – homeless people who suffer from mental health illness. Moreover, they discuss potential ways to ensure representativeness and reduce attrition – for instance, surveying individuals that use homelessness services target to particular subgroups (e.g. women, drug users, etc.) and use of technology, such as texting and social media, to maintain contact and communication with study subjects.

Moreover, there are statistical methods that can be applied to make sure that the selected sample is representative of the population of interest. Bonevski et al. (2014) discuss approaches to survey design and sampling in order to include disadvantaged groups in data collection processes. They review approaches that are used in the literature to survey populations that are hard to capture – such as non-random sampling methods (snowball), oversampling, time-location sampling, respondent-driven sampling, and targeted sampling.[11]

Existing surveys in the UK that either are explicitly designed to explore experiences of homelessness or cover homelessness among other adversities would benefit from applying appropriate techniques such as the capture-recapture method[12] to come up with samples of homeless groups of satisfactory size that are representative of populations of interest in England.

[11] For a more detailed description about the different sampling techniques and their advantages and disadvantages as well as a broader discussion about how vulnerable groups can be captured in research, see the paper by Bonevski et al. (2014).

[12] A more detailed discussion regarding the capture-recapture method is included in the “Review of models of homelessness” report.

3. Key choices for designing a homelessness model suite

This section discusses some practical issues related to designing a suite of models that address the distinct objectives around homelessness and rough sleeping. The design options discussed here apply to the set of different methods that should be included in the model suite.

Specifically, issues around the following topics are discussed:

- model development,
- ease of use,
- flexibility,
- granularity of outputs, and
- transparency.

3.1 Model development

Model development is usually a one-off event[13] that requires specialist knowledge. There is a spectrum of technical expertise and demands for resources for constructing a fit-for purpose model ranging from simple and quick to complex and time-consuming processes.

3.1.1 Internal development vs. commissioning externally

Depending on considerations about required expertise and time resources, MHCLG and DWP can choose between developing a suite of models in-house or commissioning the task externally.

In order to build a model in-house, MHCLG and/or DWP need to consider constraints related to current staff skills, staff availability (i.e. spare capacity) and software availability. Such constraints might result in a more restrictive set of development options unless specific investment is undertaken in the development of this particular suite of models.

Therefore, since model development is usually a one-off process or one that is not repeated often, it may be preferable to commission model development to a team of experts that will have access to the required set of skills.

The above discussion relates to the initial development of the suite of forecasting models and the complex simulation model, as well as subsequent development of substantial new modules.

The department should retain in-house, or invest in acquiring, the necessary skills to develop simple, ad hoc models for policy appraisal, both

[13] See the flexibility section (section 3.3) for a discussion of issues related to adapting an existing model.

to be able to provide policy analysis during the period the complex simulation model is being developed and to quickly respond to future policy analysis requests when these fall outside the current capabilities of the complex model.

Box 5. Issues around model development: software

One major consideration related to model development is the software used to build the model. There is a great range in the skill and time requirements to learn and use different software packages. Since some packages are more prevalent than others, the experience that analysts have with different software varies (e.g. MS Excel is common compared to more niche types of software).

Moreover, there is significant variation in the level of user proficiency for some software packages. A user-friendly interface is usually included in these packages to enable a large set of analysts who understand basic statistical concepts and functionality to use the tool. More sophisticated tools are also offered to specialist analysts who can perform more complex tasks such as programme their own commands and simplify model code through specialist languages. For example, the Mathematical Programming System for General Equilibrium analysis (MPSGE) language was designed to make general equilibrium modelling more straightforward to code in the General Algebraic Modelling System (GAMS) software package.

There are many options available for developing time series and simulation models, including Excel, STATA, EViews, R, SAS and Python. EViews is a statistical software primarily designed for time series oriented econometric modelling. It can be also used for the development of simulation models. STATA and EViews feature a programming environment for writing code using a consistent syntax and also include a user interface that allows users to access all available features through menus and associated dialogs.

Statistical Analysis System (SAS), R and Python are more advanced programming languages that can handle large datasets and are often used for developing complex models. Learning these languages is not an easy task and requires a good understanding of statistics and some programming skills – therefore, they are not as straightforward to use as spreadsheet programmes (such as Excel), STATA and EViews. However, analysts who have been trained to use them usually prefer them over other options as they allow for a wider range of programming functions (e.g. writing custom packages and functions). R is the most popular among statisticians and analysts as it is an open-source language and free software environment which includes a vast collection of packages, has good graphical tools as well as documentation that is available online. Overall, each software offers different functionalities. The analysts building the model will select suitable packages from the set they are familiar with or are willing to learn that best meet the model requirements.

3.2 Ease of use

A critical aspect of the developed suite of models relates to the skills and experience required to understand and apply the models to generate results. All models will take some time to understand and use, whether operated by an expert analyst or someone with less of a quantitative background.

3.2.1 Understanding the models

Some models can be understood with a few minutes of explanation and demonstration – for example, a user-friendly Excel spreadsheet where the analyst is required to select from a list of options such as the length of the forecast (e.g. 1 year, 2 year, 3 years) and the scenario (e.g. optimistic, central, pessimistic).

Others are so complex, in terms of both the structure and the computer software and code used in their programming, that being in a position to know how to run them can take weeks of training. For example, the Sub-Regional Housing Market Model (SRHMM) developed by Bramley and colleagues (see for example, Bramley, 2017; Bramley et al. 2016; Bramley and Watkins, 2016) comprises a large number of modules which quantify the effects of a broad set of interdependent factors on housing needs. The model is quite complex and was not designed to be easily used by analysts who do not have a detailed understanding of the model architecture.

In the case of such complex models, being able to independently use all model functionality can take months and might require specific domain knowledge and academic background.

3.2.2 Using the models

In addition to the time taken to understand how to use the model, there is a separate time dimension related to running the model to get results. If there are numerous forecasts to be produced and policies to be simulated, often to tight timescales, as may be the case with MHCGL and DWP's interest in homelessness, the ease of running the model is a non-negligible consideration.

Simple models can likely be programmed quickly once the analyst is familiar with model code, and running time is likely to be a matter of seconds. On the other hand, it can take time to set up different scenarios in complex models and the running time for each simulation can take hours depending on the hardware and model structure.

It should be also noted here that there is a trade-off between time and expertise needed to construct the model and ease of use. When designing a model – e.g. writing code that sets out how different variables interact to produce output – analysts need to devote additional time to make the model easy to use. Model developers can write guidance notes or programme a user-friendly front end to the model. For example, analysts from the Department of Computer and Systems Science of Stockholm University have recently developed a simulation tool for the European Commission that allows users to easily estimate the impact of changes in policy on societal outcomes based on complicated calculations that are conducted in a back-end environment.[14]

3.3 Flexibility

It is important that models used for policy purposes can be revised in order to deal with new data and/or requirements (e.g. new policies).

3.3.1 Revising and updating models

In theory, all models can be adapted to deal with new requests. However, the ease of making changes can vary substantially depending on the model's complexity. In the case where changes are difficult and time consuming, it may be preferable to build a new model rather than adapt an existing one.

For example, the version of SRHMM, which was adapted to project poverty and inequality trends (Bramley et al., 2016), includes a micro-simulation model that generates snapshots of welfare outcomes under alternative scenarios at high levels of disaggregation (e.g. for specific population segments, across localities, etc.) The development of this feature, which was added to the existing SRHMM macro-simulation model, was a complex procedure – the model was built using app. 4,000 lines of code to model the entire set of scenarios.

Models can also vary in terms of the ease of updating new data, from a simple automated procedure to more involved processes that require building extensions. Similarly, changing the set of factors that affect results can vary between typing a few letters into model code e.g. adding some variable names to an existing list, to spending months adding new components to the model which have to integrate with all existing parts in a way that is consistent with theory.

There can also be variation in how existing factors in the model affect the results. This may be required, for example, if new evidence suggests that a factor is more important in affecting an outcome than previously believed. Some changes will be as straightforward as changing one parameter in the model (this should follow a literature review to decide on the value for the parameter). On the other hand, some models may require altering the way certain elements interact with other variables, which can be a more involved process. Both options will require re-running the model and testing, as well as possibly recalibrating other factors.

[14] See here for a more detailed outline of the front-end interface that was developed to make European Commission's simulation models more user-friendly https://cordis.europa.eu/docs/projects/cnect/2/611242/080/deliverables/001-611242Sense4usD62PolicyModellingToolFINAL.pdf

It should be noted here that it may be more time-consuming to deal with model extensions in the cases where a user-friendly front end has been developed. For example, developing new code and using new data sources likely requires going back to the model developer to update the front end so that it incorporates options for the new policy options and/or data.

As a general rule, models should be developed in a manner that allows routine maintenance (e.g. inputting new data) to be done easily in-house – ideally via a suitable front end rather than requiring advanced coding knowledge or expertise in the finer aspects of how the model works. For more substantial additions (e.g. new modules allowing for additional output) such expertise will still be necessary – as it is very difficult to create a user-friendly front end enabling more fundamental changes to the model – but care should be taken to ensure that these additional modules can be developed and integrated with the main model without needing to amend the basic model architecture.

3.4 Granularity of outputs

Models developed to predict key homelessness statistics and appraise policies should be able to generate outputs at different levels of aggregation. Specifically, models should explicitly take into consideration variations in service delivery and the prevalence of different types of homelessness across English LAs, as well as the different needs of population groups who are vulnerable to experiences of homelessness.

3.4.1 Disaggregation at the LA level

Some aspects of homelessness policy operate at a national level (e.g. housing benefit), while others operate at the local authority level (e.g. homelessness prevention services). Moreover, the prevalence of different homelessness types as well as the needs of homelessness groups vary across LAs.

The development of models around homelessness should take these variations under consideration. In order to make well-informed decisions about homelessness policies, MHCLG and DWP analysts should have the ability to produce results at differing geographic levels.

3.4.2 Other levels of disaggregation

Analysts should also consider options about developing model features to generate outputs that are granular at levels other than geography. Homelessness interventions are often explicitly designed to support households and single people with complex needs who are more vulnerable to adversities such as homelessness. Therefore, the suite of models around homelessness should be able to produce projections of homelessness and appraise the effects of policy changes for specific subpopulation groups – e.g. BME, people who suffer from mental illness or substance abuse, young people who have left care.

To arrive at granular homelessness projections, models fitted to aggregated national data including indicators for LAs and other socio-economic characteristics of interest (e.g. gender, age, institutional history, etc.) can be used. This option is optimal when survey sample sizes or collection of administrative data at higher levels of disaggregation are too small to produce outputs specifically for each group or area of interest. Alternatively, models can be separately applied to LA-specific series of data or sub-samples with specific characteristics. For example, the forthcoming collections of administrative data on homelessness (H-CLIC), which include background information of applicants for LA homelessness services, can potentially allow for estimating models using series of data that are specific to each LA or particular population segments across LAs.

3.5 Transparency

There are two dimensions to model transparency:

- whether there is a published detailed description of the model, including model code, and
- whether the outputs of the model are published/shared.

3.5.1 Transparency of code

Open source models are usually based on code that is shared with the public. They represent the most transparent modelling option. At the other end of the spectrum are proprietary models that are not usually available to the public.

There are pros and cons of varying the transparency of both model description and outputs. Open source models often have more user-generated guidance discussing how to overcome various problems/frequently asked questions than non-open source models, especially when using less well-known software packages. One frequent criticism with proprietary models is that their results can be a ‘*black box*’ – that is, it is difficult to understand *how* results were obtained. It is good practice to publish robustness tests and sensitivity analysis to deal with the black box accusation, but in practice this rarely occurs outside of academic research.

In some cases of models that are important to the public, some model elements are made publicly available. For example, the public has access to a set of core equations from the Bank of England Macroeconomic Model (see, for example, Hendry and Muellbauer, 2018). Moreover, detailed revisions of model equations are published regularly from the Office for Budget Responsibility (OBR, 2013).[15]

In practice, even in these cases

[15] Available here: https://obr.uk/docs/dlm_uploads/Final_Model_Documentation.pdf

members of the public cannot run the model based on the publicly available information. Specialist knowledge is required to actually use the models - e.g. refining output to improve results or changing key parameter values which are often not published.

Our assumption is that the homelessness models will be mostly proprietary – for exclusive use by MHCLG and DWP. These models will likely be used to inform confidential advice about policy development within the two departments. However, there may be elements of homelessness models that could be made open source either once models are unveiled or at a later date. Having open source components would allow academics to review these aspects of the model and suggest improvements or extensions.

3.5.2 Transparency of outputs

In addition to transparency around model code, we also consider transparency around model outputs. While evidence feeding into policy appraisal is almost always unpublished, there may be cases where homelessness modelling output is securely shared with a select group. For example, it may be useful to share outputs predicting homelessness at the LA level in the short term (one to three years) with LA teams who can use the results to plan for their homelessness services, such as temporary accommodation provision. It could also serve as a useful sense check of the forecasts.

4. Considerations of model specific issues

4.1 Time-series models

Time-series models are frequently used to produce short-term forecasts. In forecasting, short-term usually refers to a period of up three years in the future. There are time-series models that only include historic values of the series being forecast (e.g. predicting family homelessness in the UK next month using only historic values of family homelessness) called univariate time-series models. These univariate time-series models include:

- models that use the tendency of a series to return to its path following shocks (ARIMA modelling),
- models that estimate a series' tendency to behave differently over different periods (regime-switching models), and
- models that forecast a series' volatility (GARCH).

There are also multi-variate time-series models which can include historic values of the series being forecast but also integrate other variables to predict a series, such as vector autoregression (VAR) models or error correction models (ECM). In order to arrive at reliable forecasts of homelessness, model developers should explore the scope for including a limited set of indicators that reflect key determinants of homelessness (e.g. poverty, affordability) in multivariate time series models.

These models may not be structural in the sense that they do not generally attempt to identify factors that cause homelessness or the process through which various factors interact to result in an individual or household becoming homeless. They should be assessed in terms of one quality – forecast accuracy – that is, their success in prediction.

Forecast accuracy can be measured in a number of ways. A standard technique to assess forecasts is to compare the mean squared error (MSE) of out-of-sample forecasts.[16] As the name describes, the mean squared error is the average of the square of the error of the forecast – that is, the difference between the predicted value and actual value – over all time periods available. This measure places equal weights on positive and negative errors, and errors over time. While MSE is frequently used in forecasting, there are many alternative techniques that can be used to assess forecast accuracy.

[16] In a nutshell, forecasters perform out-of-sample validation to test the accuracy of time series models. This means that they first exclude some data from the sample when identifying and estimating the model and then use the model to predict this out-of-sample figures. Forecasting accuracy is assessed by comparing forecasts with actual out-of-sample data as well as forecasting errors with the errors of the model when fitted in the selected sample.

4.1.1 What questions can the model answer?

Time-series models answer the set of questions that ask:

"*what is the value of a variable in the near future?*",

where near future can be any time period up to three years ahead.

Even if the time-series model contains explanatory variables that have been shown in the literature to cause homelessness, analysts should refrain from the temptation to draw policy implications from the results. The models are essentially built to provide accurate predictions of short-term trends in outcomes of interest, rather than determine links of causality between covariates and outcomes of interest.

4.1.2 Data

Sequential sets of data points measured at frequent time intervals are used to estimate time-series equations. Historical data reported at various levels of frequencies (e.g. monthly or quarterly) are often used to gauge future trends in outcomes of interest based on the assumption that existing trends will continue to take place in the short-term. The data inputs are of primary importance when determining the frequency of forecasts (e.g. monthly, quarterly or annually). Unsurprisingly, higher-frequency data will result in more accurate forecasts that take into consideration potential seasonality in the outcomes of interest.[17]

Moreover, time-series models can include a limited set of explanatory factors (multivariate models). Therefore, the level of detail and granularity in collected data will influence the level of disaggregation in the forecasts (e.g. in terms of geography, homelessness forms and individual characteristics).

In order to generate short-term forecasts of homelessness, series of frequent administrative[18] data such as local authority counts of homeless and rough sleeping groups should be used. At present, the main source of data on statutory homelessness is official counts of people in temporary accommodation collected by the LAs and reported by MHCLG on a quarterly basis. These series of data also include information on the broad characteristics of households owed a statutory homelessness duty, including those who are housed in temporary accommodation. Such data can be used to produce estimates of future trends in statutory homelessness. Another example of low frequency administrative data that can be used

[17] While we are not aware of any evidence pointing to seasonal variations in homelessness, research overseas suggests that there might be seasonal changes in forms of homelessness and use of support services such as emergency accommodation (see, for example, Colburn, 2017)

[18] In practice, most of the data series used for forecasting are likely to come from administrative sources though there might also be the potential to use data from other sources (e.g. long-running surveys).

to generate short-term forecasts are the annual counts and estimates of rough sleeping populations reported by LAs.

Recent data collections, such as the H-CLIC series and experimental data on statutory homelessness, that will go beyond prior collections by covering other forms of homelessness (e.g. single people homelessness) will enable the production of forecasts for types of homelessness other than statutory homeless households in priority need. Moreover, the dataset includes information about demographic characteristics and other complex needs, potentially allowing for further disaggregation and segmentation of homelessness forecasts.

4.1.3 Model development

In general, simple time-series forecasting models can be constructed relatively quickly. However, detailed research into developing and refining accurate forecasts would be a more substantial project.

Time-series modelling is conducted by testing a number of available specifications to arrive at the method (or the combination of methods) that result in the most accurate forecasts. This process is likely to require a number of iterations. The number of specifications to test – in terms of type of time-series model and set of factors to include – increases exponentially as the number of series that need forecasting grows.

If there is a fairly contained number of series required, e.g. up to 100, it would be relatively straightforward to calculate each one, so the set could be completed relative quickly. A group of experienced forecasters could have an initial set of results for a long list of series of interest within two-three months, and further develop forecasts within six months.

If the number of series that need forecasts is prohibitively large to allow analysts to design tailored models for each series, there are options for using systematic rules to produce groups of forecasts. For instance, if forecasts are needed for three types of homelessness in every LA with 20 different population segments (with specific characteristics), models should be developed to generate over 9,000 forecasts.

The application of rules to automate the forecast process (e.g. all forecasts for family homelessness in coastal and rural areas follow the same model) facilitates the generation of a large number of outputs. A smaller set of the most useful forecasts could be then updated by analysts.

Finally, the most important requirement for developing a time-series empirical strategy is knowledge of forecasting. This is because time-series models primarily require technical time-series forecasting skills rather than sector-specific knowledge.

4.1.4 Ease of use

Ideally, the team of analysts that will run the time-series models should have an understanding of homelessness and rough sleeping. This way, they will be able to sense-check model outputs, explain potential variations in results that cannot be accounted for by determinants and be aware of any structural breaks.

In contrast to complex policy models – where developing a user-friendly front end is an involved undertaking – developing an easy to use, intuitive front end for a series of forecasting models should be relatively easy to do at model development stage. Once the models have been developed, running the forecasts regularly, e.g. every quarter or month, should be straightforward to do for an entry level analyst. It should also be relatively easy to feed in new data as they become available. Assuming a front end has been built, the models could even be made accessible to non-analysts with basic numeracy skills.

4.1.5 Flexibility

While the frequency of updating and revising time series models depends heavily on the setting and the series being modelled, it is critical that they are updated as soon as new data is made available. Moreover, the model parameters should be re-estimated every few years to make sure that the model generates accurate forecasts. Finally, forecasters should carry out a more fundamental development of the model in five to ten years to ensure that up-to-date techniques are implemented to maximise forecasting accuracy.

Updating forecasting models with new data is a straightforward task. However, users should keep in mind that the models will need to be revised every few years to allow for potential changes in the parameters. Revising the models requires a team of experts that have advanced technical skills and are familiar with forecasting techniques.

In terms of revising the series to respond to new policy objectives and interventions (e.g. new target groups of policy interventions), time-series models present vast options for flexibility as they can forecast any type of series. If there is sufficient data available in terms of how many individuals are included in the specific forecast, there can be any combination or combinations of different geographic levels, individual characteristics and types of homelessness. Where numbers are too small to allow analysts to build time-series models (e.g. instances at a very granular level), there are alternative options for forecasting.

Time-series models are also flexible in that they can integrate expert judgement as well as administrative or survey data. For example, if there was a survey asking experts working with homeless people in LA teams whether they believe homelessness will increase, decrease or stay the same over the next quarter, their responses could be included in the forecast through an additional modelling process. Analysts can verify whether adding a judgement variable improves forecast accuracy.

4.1.6 Granularity of outputs

In the context of homelessness, there may be a whole range of series for which government analysts would like to produce short-term forecasts. These can span aggregate figures for the overall homeless population, to smaller groups such as figures for each type of homelessness, to much more disaggregated specific groups (e.g. number of new members of the homeless population under age 25 living in the West Midlands).

One strength of time-series models is that they can be very flexible in terms of disaggregated results, depending on data availability. In terms of geographic disaggregation, one can either produce a national level forecast or calculate forecasts for each LA and aggregate up to produce a national-level series. Selecting between these options will depend on whether one believes that each LA should have different parameters included in the forecast models – a choice that can largely be tested during model development. These models could also be produced at the level of groups within LAs (e.g. split between urban and rural areas).

The frequency of any short-term forecasts is driven by data availability and analysts' judgement. For instance, data inputs that are updated at least once per quarter (or at higher frequencies – e.g. monthly) are required to generate quarterly forecasts of a certain series.

4.1.7 Transparency

Simple time-series models do not generally offer a direct causal interpretation of coefficients and it can therefore be difficult to communicate the meaning of model equations. These models predict what will happen, not why (Hyndman and Athanasopoulos, 2012).

However, since the models are usually short equations, it is straightforward to publish equations and estimated coefficients. The relative simplicity of these models lends them some merit, as scrutinisers with a general economics background would largely be able to understand and test the assumptions.

4.2 Simulation models

Simulation models are more diverse in terms of the techniques involved and their general uses. When thinking about homelessness modelling, there will be two broad classes of interest:

- simple (ad hoc) models that appraise a limited set of policies – developed in short time frames to provide quick advice for planned policies, and
- more complex models that predict long-term projections of future trends – setting a baseline and/or assuming composite changes in economic and policy variables.

Simple ad hoc models are usually designed to appraise a specific policy (or policies). They estimate the additional effects of policy changes while remaining agnostic to baseline trends in homelessness, which are considered given. Therefore, selecting between simple and more complex simulation models mainly depends on available resources, research questions and requested outputs.

On the other hand, more complex simulation models arrive at long-term projections by taking into consideration the effects of a broad set of complex relationships between personal, economic and policy factors that predict homelessness. These models can be used to produce reliable long-term projections assuming no changes in the economic and policy environment. Moreover, they can be applied to estimate outcomes under composite policy and economic scenarios. It follows that the main difference between simple and complex models is that the latter are comprehensive – they model homelessness as the outcome of complex links and interdependencies between a wide set of predictive factors.

The model developed by Bramley and colleagues is an example of a complex model that accommodates both objectives – policy appraisal and projections of long-term trends.[19] Another example is the Intra-Governmental Tax and Benefit Model (IGOTM) which simulates the income distribution[20] in a steady-state and then estimates the impact of changes to tax and benefits on household income. A number of policy changes can be incorporated in these example models to identify the overall impact of a package of policies. If policy changes are modelled separately (adding separate components to quantify the impact chain of introducing each policy), then the distinct impact of each individual policy can be demonstrated.

Finally, simulation models (both complex and ad hoc) can allow the assessment of the Value for Money (VfM) of planned policies. In order to calculate VfM assessments, the model must link financial indicators (for policy costs and expenses) to the chains of impacts that simple ad hoc models quantify. This way, simple policy models can measure monetised impacts of planned policies. For example, models can estimate the impact of £1 spent in increased housing benefits on temporary accommodation costs – e.g. £1 spent in increased housing benefits leads to £x reduction in temporary accommodation costs due to a reduction in the number of homeless households entitled to homelessness duties.

[19] See the “Review of models of homelessness” report for a detailed description of the housing needs models developed by Bramley and colleagues.

[20] From a combination of the Living Costs and Food Survey (LCF) and The Effects of Taxes & Benefits on Household Income (ETB), which provide information on income, expenditure and important family characteristics. See Tonkin and Stoyanova (ONS 2015) available here: http://webarchive.nationalarchives.gov.uk/20160106064101/http://www.ons.gov.uk/ons/dcp171766_409063.pdf

4.2.1 What questions can the model answer

Models that set a baseline can answer questions such as:

"What are the trends for homelessness over the next 5-10 years?".

Models intended for appraisal of specific policies[21] will answer questions such as:

"What is the (likely) impact on homelessness from changes in:

- *levels of welfare benefits and eligibility for support for housing costs (including Universal Credit and Housing Benefit),*
- *housing supply (including affordable housing and policies such as the Right to Buy programme), and*
- *population in-flows (from immigration, residential care, prison release)".*

4.2.2 Data

Simulation models generate projections of outcomes of interest conditional on a broad set of predictors. The models that will be developed to project future trends in homelessness under the existing policy framework or alternative policy scenarios will include a number of factors that are shown to predict homelessness.

Evidence suggests that homelessness is a complex phenomenon, triggered by a broad set of relationships between a multitude of factors and circumstances rather than the outcome of a single event (domestic abuse) or personal characteristics (e.g. suffering from mental health problems). These include both structural factors, including house prices and policy variables (e.g. housing benefits), as well as personal factors, such as relationship breakdown and financial strain. Furthermore, the interactions between factors that are identified as homelessness causes (poverty, domestic violence, mental illness) have important effects on homelessness.[22]

The set of covariates that will be included in the model depends on its objectives and level of complexity. For example, ad hoc models developed to appraise particular homelessness interventions seek to quantify the impact of launching new policies compared to a given baseline and are not designed to consider the complex mechanisms that determine baseline homelessness levels. Such models accommodate the policy appraisal objective by including a limited set of factors that are relevant to the particular intervention. On the other hand, a more complex model aiming to project homelessness levels in the longer-term usually includes a broader set of variables to arrive at reliable

[21] We use the term 'policy' in a wide sense to also include potential issues pertaining to administration.

[22] For a more detailed discussion about evidence on causes of homelessness, see the separate report under the title "Rapid Evidence Assessment".

estimates of future trends in homelessness.

Naturally, the more factors that are explicitly included in a model (either baseline or policy simulation), the more data will be required. For example, for a model to consider factors relating to all potential causes of homelessness and the links between them, data on poverty, demographics, health and housing markets are required. Overall, even a relatively simple simulation model will be more data-intensive than time-series forecasting models.

Simulation models are also more sensitive to data quality – for example, a very complex and data-intensive model might be more reliant on smaller surveys (in absence of data of better quality) that may be subject to problems such as measurement error. Unsurprisingly, better quality of data, such as higher level of granularity and more detailed background information on applicants for homelessness duties, results in more reliable outputs.

Different sources of data can be used, depending on the relationships between covariates and outcomes of interest that each model aims to capture. Simulation models often use a combination of survey and administrative data to quantify the links between a selected set of explanatory variables and outcomes of interest. For example, UKHLS data, data from smaller scale datasets (e.g. PSE), administrative data on homelessness, official statistics on local housing and labour markets and national economic indicators can all feed into simulation models.

As discussed in the review of models around homelessness, simulation models comprise two stages:

i. links between explanatory variables and outcomes of interest are quantified, and

ii. these quantified effects are applied to projected changes in explanatory factors to arrive at projections for outcomes of interest.

Individual-level data are strongly preferred even for models looking at national policy changes, especially in the first stage, since it will increase the flexibility of the model – for example, by capturing behavioural responses to a set of changes in housing markets and personal circumstances. Even at the second stage where simulations are conducted, individual level data are preferred as they increase granularity of outputs. A representative example is the micro-simulation model included in the SRHMM macro-simulation model to project poverty and inequality outcomes at the household/individual level using data from Understanding Society (UKHLS).[23]

Linked administrative data, as well as household- or individual-level survey data, can be used to quantify the elasticity of homelessness to changes in predictive factors. In this case, the reliability of outputs will depend on the scale of the survey (large scale vs. small scale), the representativeness of

[23] For a more detailed description of economics-based simulation models (e.g. first stage/second stage) see the "Review of models of homelessness" report.

the sample and the accuracy of reported information (which might be subject to self-reporting and memory bias). Moreover, survey data can be used alongside administrative data to capture the shares of homeless populations that are hidden (for example, not being included in homelessness registries because they do not turn to LA services for advice and support).

Box 6. Causal vs. predictive factors

As outlined above, when building models that forecast medium to long-term trends, we can include predictive factors, causal factors, or a combination of the two.

Causal factors are those that directly lead to a change in the outcome of interest – if we actively intervene to increase a causal factor, we can be certain that the outcome variable will change. This kind of factor is often based on a theoretical relationship. Causal models are those which attempt to estimate how a set of factors (often referred to as explanatory variables) affect the outcome of interest.

At first glance, it may seem that causal factors would be preferred over predictive factors in a model that predicts future trends. However, causal factors may not be as useful, e.g. if they are only weakly related to the outcome variable. In this case, we should include predictive factors such as housing prices, supply of affordable housing, or receipt of welfare benefits in our model because they would allow us to more accurately predict which individuals are likely to face homelessness in the future.

Predictive factors, on the other hand, are only based on relationships we have observed in past data. There is not necessarily a mechanism that leads a predictive factor to cause the outcome of interest, but the factor does contain useful information that helps us identify future values of our outcome variable.

Essentially, if our sole aim is to arrive at more reliable predictions of homelessness rather than understand the underlying mechanisms that drive homelessness, predictive factors may be more suitable compared to causal factors with weak effects on the outcomes of interest.

To highlight this distinction, consider a research study testing whether people who experience relationship breakdown are more likely to become homeless compared to people with no such experiences. If we observe in our research population that more people with relationship dissolution experiences end up homeless, this means that relationship breakdown is a predictive factor for homelessness – however, this is not enough evidence to decide whether relationship dissolution is a causal factor. Only once we identify a mechanism being present in people who have gone through relationship breakdown but absent to people without such experiences which triggers homelessness, we can classify relationship dissolution as a causal factor.

4.2.3 Model development

As set out earlier, simulation models can be either complex, when they attempt to explicitly account for many different causal aspects, or relatively simple (i.e. more ad hoc), when they focus on limited factors or policies. Whether the model is built to establish a baseline, appraise composite policy scenarios, or evaluate specific planned policies, there can be a range of complexity in model options, depending on how many relationships are explicitly captured in the model. Models can include components that explicitly quantify all relationships of interest, as is the case with the models developed by Bramley and colleagues. In such models, the first stage estimates a set of relationships that are then used as inputs into the second (simulation) phase of the model.

There is also an option to use existing published literature to inform a relationship from the UK or abroad. For example, published research reports findings on the quantified impact of mental health on homelessness – assume that those with mental illness are shown to be *X1-X2* times more likely to be homeless. In this case, analysts would use a number within the *X1-X2* range to model the relationship between mental health and homelessness. However, the REA, which was conducted as part of the wider project on causes of homelessness and rough sleeping, revealed that evidence on the quantified effects of a set of important drivers on homelessness is rather limited.

Moreover, simulation models offer a set of options that allow analysts to adapt the selected methods to policy objectives and theoretical assumptions about the relationships they want to capture and quantify. Choosing between options related to the following issues is considered central to conducting homelessness simulations:

- *modelling level* – analysts can choose between micro-simulations that model outcomes of interest at the individual and household level and macro-simulations that produce outcomes at higher levels of aggregation (e.g. homelessness projections at the national level). A more disaggregated level model offers more flexibility.[24]
- *dynamic relationships* – modelling dynamics are important for a topic like homelessness, as evidence suggests that factors interact over time to determine individual pathways in and out of homelessness. Analysts should decide the extent to which dynamics are captured (e.g. will the model consider how interactions evolve over time? Will the model incorporate behavioural responses that may vary over time?)

[24] See the flexibility section (4.2.5) for more information

Box 7. Issues around development of complex simulation models: modularity

It is possible to build a large model in separate components – known as modular model building. Modular models can be developed in stages – with the option for separate modules being developed by different teams or different organisations.

This may be useful as large-scale modelling can take time and entails risks (for example, analysts with expertise can leave organisations, possibly jeopardising completion of model development). Breaking down the modelling in smaller modules can minimise the risk of losing specialist skills, especially when models are built using complex software and code.

There are two key points to keep in mind in implementing a modular approach. Firstly, it is critical to consider all modules that will be needed at the start of model building – even if there are no short-term plans to extend the model, it is important to plan the wider system architecture up front, taking into account what future needs might be. Secondly, it is important that modules are designed and built in a consistent manner – e.g. producing consistent (therefore, comparable) outputs, using the same software, following the same naming conventions etc.

These good practice examples apply when adding new modules in the future and adjusting or improving existing modules. For example, if a behavioural element, such as how individuals respond to a particular policy, should change, a model where the code to make such an adjustment is in as few places as possible (to avoid making large, difficult structural changes) is preferred. It follows that planning for future adjustments at the model development phase can save time.

4.2.4 Ease of use

Using a model will be less time consuming if resources are allocated to the development of a user-friendly front end. This may be more relevant for simulation models that are very large and complex. Using tools such as drop-down menus would allow users to manipulate key parameters (e.g. select geographic levels and population segments) to produce disaggregated outputs or changes in trends as a result of changes in policies without the users having to understand how the simulations are coded.

It is important to develop a front end to the model to allow in-house analysts with solid analytical background to apply the model without first needing to be familiar with the complex underlying model structure.

4.2.5 Flexibility

Ad hoc models are usually developed quickly without considering flexibility in order to answer specific policy questions. While such models can be relevant for longer time periods, they can generally be seen as one-off exercises serving a particular purpose. In this context, updating and revising the models often exceeds the purpose of their development.

On the other hand, when designed to be modular and flexible, the basic structure of more complex simulation models can potentially remain relevant for decades (e.g. Institute for Fiscal Studies (IFS) TAXBEN model,[25] HM Treasury and OBR Macroeconomic model[26]). This does not mean that the models do not require maintenance. The models should be regularly updated using new data as soon as they are published. It is critical that this task is carried out by non-expert analysts. Moreover, certain parameters should be re-estimated every few years to make sure that the model results in reliable projections that reflect actual homelessness trends. Finally, the models should be fully revised when fundamental changes in the policy and economic environment take place (e.g. change in tax system, transition to Universal Credit, etc.)

A more complex simulation model will include many distinct model elements to allow for interconnected relationships between a wide range of factors that influence outcomes of interest. For example, models can potentially include a range of micro and macro factors such as poverty, demographics, benefits, health, social care, housing markets and labour markets.

In general, the more components are included in a model, the more flexible a model can be in terms of dealing with policy changes and capturing all relevant mechanisms of effects transmission. That said, it may be less straightforward to adapt a more complex model to deal with new extensions compared to a simpler ad hoc model.

---

[25] See here for more information https://www.ifs.org.uk/publications/572

[26] See here for more information https://obr.uk/docs/dlm_uploads/Final_Model_Documentation.pdf

A way to avoid excessive complexity that hinders revising and updating the models is to adopt a modular model architecture (see box 8 for a more detailed discussion). A modular simulation model on types of homelessness and rough sleeping will comprise a broad set of separate modules to estimate future trends in each predictor (e.g. poverty, affordability, housing supply, unemployment, welfare benefits, key economic indicators). These modules should be self-contained and result in clear outputs that will then feed into the core simulation function that produces homelessness projections.

The key point here is that consideration of future needs for potential additions of model components facilitates model revisions at later stages. Moreover, it is important that a complex model be simplified as much as possible – for example, interactions between different variables in a model can be avoided as they are complex and difficult to interpret without adding to the statistical power of a linear model.

Complex models on homelessness are likely to include a number of inter-related modules producing outputs that are then used as inputs in other stages of the model. For example, a simulation model can include separate modules to estimate future levels of affordability and housing supply conditional on a set of predictors. Housing supply predicts affordability outcomes – therefore, outputs from the housing supply module will feed in the affordability model as inputs.

Estimates of housing supply and affordability can also be used as inputs in the core simulation model that produces homelessness projections. This process of inserting non-static outputs from particular modules into other modules as inputs is likely to result in increased complexity. Therefore, an important consideration when developing a flexible simulation model is to make sure that users are able to clearly track the links between its different elements.

Model flexibility in terms of output types and disaggregation depends on available data and modelling level (e.g. micro vs. macro-simulation). For example, models that are based on individual data will simulate behaviours for every individual that can then be aggregated to any group of interest.

4.2.6 Granularity of outputs

Since simulation models are often built up from administrative or survey data at the individual level, it should be straightforward to produce granular outputs depending on non-disclosure requirements.[27] Disaggregated survey and administrative data can be used as inputs to models that aim to estimate key predictors of homelessness at the local level or for specific subgroups. Outputs from these modules at the first stage of a simulation model (rather than actual survey or administrative data) will be then inserted in the core functions of the simulation model to arrive at projections of homelessness and

[27] Models can, and should, be adapted to automatically suppress potentially disclosive output.

rough sleeping.

In addition to producing outputs disaggregated at lower geographical levels (e.g. across LAs), simulation models can generate projections of homelessness that are specific to groups of policy interest. Depending on policy aims, they can be used to predict levels of homelessness among specific target groups – for example, people aged 18-19 with complex needs who have recently left care or people with substance dependence problems who return to homelessness and rough sleeping prevention services for support and advice.

There are a number of options to build simulation models that produce outputs at various levels of disaggregation. These options range from including indicators for personal characteristics of interest (e.g. gender, age, whether suffering from mental illness, have left care, etc.) and level of geography (e.g. LA indicators) to developing bolt-on modules that allow for further disaggregation of outputs across groups of interest.

An option for producing outputs that can be further broken down by locality or by specific population group is the development of micro-simulation models. These models generate projections at the household/individual level conditional on projected long-term trends in demographic, policy and economic variables. There are a number of options for the development of a micro-simulation model – for example, model developers can choose between dynamic and static micro-simulation. Dynamic micro-simulation models are quite complex and are often hard to revise and update. Moreover, some forms of micro-simulation may be incompatible with any kind of sub-regional disaggregation.

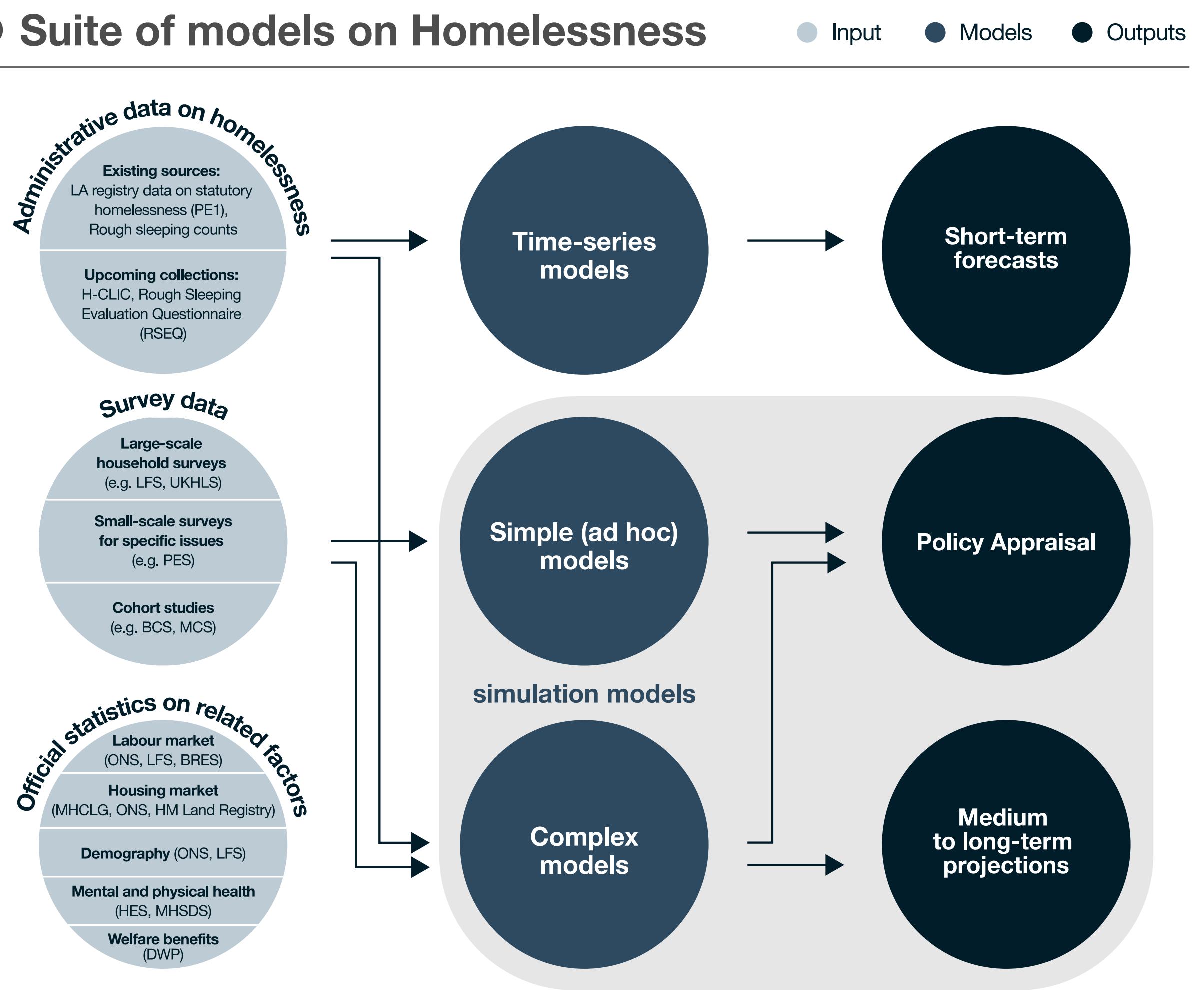
4.2.7 Transparency

When considering the spectrum from simpler to more complex simulation models, simpler models tend to be more transparent in terms of explaining how results are produced (including examination of model code) and interpreting these results.

Large-scale and complex models can be more difficult to explain. The complexity of interactions and wealth of parameters included (especially if there are multiple stages to model estimation) means there are risks of the model and its results becoming a 'black box' where complete understanding is only available to specialist analysts. Caldara et al (2012)[28] suggest that as layers of complexity and interaction are added, the results become more opaque and harder to explain to policymakers.

[28] Caldara et al. (2012) discuss issues related to the development of Dynamic Stochastic General Equilibrium models, though their point is true for complex simulation models more generally.

Suite of models on Homelessness



Moreover, model complexity also requires more advanced scrutiny. The more complex a model is, the more it will require analysts to make judgements during development and generation of outputs – judgements can change results, sometimes significantly. The more complex a model is, the more likely that scrutiny (either internal or external) would require specialist training.

There are ways to enhance the transparency of simulation models irrespective of whether they are simple or complex. Simulation models can be intuitive to communicate if modelled relationships are well-documented and underlying equations are based on established theory.

However, even in the cases where quantified links are transparent and depend on substantial assumptions, there can be complications to explaining a simulation model – for example, how weights are applied to scale results at the national level, how certain parameters have been assumed/estimated, where analytical judgement is applied, etc.

Finally, it is critical that documentation, notes to the code and guidelines for use are produced at the development stage of a complex simulation model to allow for analysts with different backgrounds to understand and apply the model.

5. Conclusions and recommendations

The purpose of this scoping exercise is to present MHCLG and DWP with a set of options for developing models to address the following distinct objectives around homelessness:

i. short-term forecasts,
ii. medium to long-term projections of future trends, and iii. policy appraisal.

Recommendations for a suite of models

Our recommendations for model development are guided by the key findings from reviewing and assessing the suitability of existing models on homelessness[29] – particularly, that using different models to address different objectives is preferred to using a complex model to accommodate all desired objectives. Moreover, when identifying available options, we considered the need for relatively quick model development to inform policy in the short-term as well as the need for a more comprehensive model that would produce more detailed outputs on a consistent basis.

Based on these considerations, we suggest that a suite of models comprising the following elements be developed to address the three policy objectives:

- time-series models for accurate short-term forecasts,
- simple, ad hoc simulation models for appraisals of specific policies, and
- complex simulation models for medium to long-term projections as well as policy appraisal at a more detailed level.

Considerations about data inputs

All the suggested models can be applied using existing sources of data. However, data quality and availability influence model outputs – more detailed data lead to more reliable outputs under the same empirical design. Overall, improving data collections on homelessness and rough sleeping is an important task that will contribute to a more robust estimation of homelessness levels and policy appraisal.

The following recommendations to enhance homelessness data collections will contribute to more reliable outputs and, thus, well-informed policy decisions:

- covering all homelessness types – it is important that LAs start to gather and frequently report detailed information on types of homelessness other than statutory homeless households and people rough sleeping,
- data linking – linkage of data between various administrative sources allows for capturing relationships between a wide set of predictive factors and

[29] For a detailed review and assessment of existing models, see the "Review of Homelessness Models" report.

homelessness outcomes, and

- consistency and data sharing across LAs – good communication and collaboration between LAs will result in more consistent and accurate counts of homeless populations as well as more reliable reporting of related factors (for example, support needs). MHCLG and the Local Government Association can play a key role in enabling consistency through developing a common framework for data collection and disseminating best practices. The local data pilots that will be launched soon as part of the Rough Sleeping Strategy are an example of an initiative undertaken by MHCLG to develop standards for consistent data collection across LAs.[30]

Important modelling choices

A number of options around development, ease of use, flexibility, granularity of outputs and transparency should be considered. The models that will be developed should:

- be easy to use – resources should be invested to develop a front end that allows users to work with even complex models without having to be familiar with how core equations are coded,
- be straightforward to update and revise using in-house expertise – the departments should invest in training in-house analysts for model maintenance, operation and periodic updates, and
- produce outputs at high levels of granularity – given that policy decisions on homelessness are taken at the LA level and may focus on particular population segments (e.g. people with drug abuse issues), it is important that models generate homelessness estimates at lower levels of geography, different segments of the population and combinations of the two.

In addition to these general considerations, strategies for developing each model type recommended for the model suite need to consider options around these issues that are relevant to model-specific characteristics.

Time series models

While time series models are simple equations that are relatively easy to use once finalised, developing a flexible set of models requires a team of analysts with strong technical skills and expertise in forecasting techniques. Therefore, MHCLG and DWP may wish to explore the scope of externally commissioning the development of a set of forecasting models that can use a large number of series to arrive at granular forecasts of different types of homelessness and rough sleeping. Resources should be allocated to the development of a front end that allows in-house analysts to use the models to produce forecasts and scenarios as well as feed in new

[30] See here for more information on MHCLG's Rough Sleeping Strategy https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733421/Rough-Sleeping-Strategy_WEB.pdf

data as they become available.

Simulation models

Building an all-encompassing simulation model that produces medium to long-term projections and appraises composite policy and economic scenarios using a wide range of predictive factors should be the key aim, as this model can offer key insights into homelessness and enable effective policy decisions to address problems. The development of the model could be commissioned externally, but it should be possible to fully operate and maintain the model (including data updates) in-house. MHCLG and DWP should play an active role during the model development process both in terms of clearly specifying the brief and in terms of interacting with the model developers to ensure key design elements are fit for purpose.

Given the time required to develop this complex model, it is important the departments retain, or invest in acquiring, in-house expertise to build and use simple ad hoc models for policy appraisal. This capability should be retained even following development of the complex model discussed in this report to sense-check its outputs and incorporate future requests for ad hoc analysis not accommodated by the complex model.

A critical aspect to the development of a complex model is the design of a front end that allows users to operate the model easily. It is important to invest in producing documentation and guidance that allows in-house users to operate the models and carry out routine maintenance. Since making complex policy models user-friendly can entail significant costs, the optimal solution is to develop a front end that can be used by in-house analysts with the relevant academic background and some limited, model-specific training, but not necessarily aim for a genuinely ‘consumer-grade’ level of user-friendliness.

Finally, it is critical that the complex model has a modular structure, comprising of a large set of modules to model complex links between predictive variables and homelessness outcomes that can be developed and adapted independently of each other. It is important that model development allows for different elements to be developed by different teams of analysts and that it is possible to adapt the model or add new modules with minimal need to alter the basic model architecture.

References

Aspinall, P.J., 2014. Hidden needs: identifying key vulnerable groups in data collections: vulnerable migrants, gypsies and travellers. University of Kent, Centre for Health Services Studies.

Beatty, C., Cole, I., Powell, R., Kemp, P., Brewer, M., Emmerson, C., Hood, A., Joyce, R., 2014. The impact of recent reforms to Local Housing Allowances: summary of key findings. Research report no 87. Department for Work and Pensions.

Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., Brozek, I., Hughes, C., 2014. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Medical Research Methodology 14, 42. https://doi.org/10.1186/1471-2288-14-42.

Bramley, G., 2017. Homelessness projections: core homelessness in Great Britain. Crisis, London.

Bramley, G., Leishman, C.M., Cosgrove, P.J., Watkins, D., 2016. What would make a difference? modelling policy scenarios for tackling poverty in the UK. Heriot-Watt University.

Bramley, G., Watkins, D., 2016. Housebuilding, demographic change and affordability as outcomes of local planning decisions: Exploring interactions using a sub-regional model of housing markets in England. Progress in Planning, 104, 1-35.

Brewer, M., Emmerson, C., Hood, A., Joyce R., 2014. Econometric analysis of the impacts of Local Housing Allowance reforms on existing claimants. Department for Work and Pensions, Research report, no. 871. DWP, London, UK. Available at: https://www.bl.uk/collection-items/econometric-analysis-of-the-impacts-of-local-housing-allowance-reforms-on-existing-claimants

Colburn, G. 2017. Seasonal variation in family homeless shelter usage. Housing Policy Debate, 27 (1), 80-97. https://doi.org/10.1080/10511482.2016.1158200

Fitzpatrick, S., Bramley, G., Johnsen, S. 2012. Multiple exclusion homelessness in the UK – An overview of key findings. Briefing paper No1. Heriot Watt University.

Fitzpatrick, S., Bramley, G., Sosenko, F., Blenkinsopp J., Wood, J., Johnsen, S., Littlewood, M., Watts, B. 2018. Destitution in the UK 2018. Joseph Roundtree Foundation.

Fitzpatrick, S., Pawson, H., Bramley, G., Wilcox, S., Watts, B., Wood, J. 2018. The homelessness monitor 2018. Crisis, London, UK.

Gordon, D., 2011. Main PSE UK sampling frame. Working Paper Method Series No.21. Poverty and Social Exclusion Survey.

Hendry, D. F., & Muellbauer, J. N. (2018). The future of macroeconomics: macro theory and models at the Bank of England. Oxford Review of Economic Policy, 34(1-2), 287-328.

Hough, R. L., Tarke, H., Renker, V., Shields, P., Glatstein, J. 1996. Recruitment and retention of homeless mentally ill participants in research. Journal of Consulting and Clinical Psychology, 64(5), 881.

Hyndman, R.J., Athanasopoulos, G. 2012. Forecasting: principles and practice. OTexts. Available at: https://www.otexts.org/fpp

Kelley, K., Clark, B., Brown, V., Sitzia, J., 2003. Good practice in the conduct and reporting of survey research. International Journal for Quality in Health Care 15, 261–266. https://doi.org/10.1093/intqhc/mzg031

Maher, J., Drever, E. 2013. Poverty and Social Exclusion Survey technical report. UK Data Archive Study Number 7879. The National Centre for Social Research.

Ministry of Housing, Communities and Local Government, 2018. National evaluation of the Troubled Families programme 2015-2020: family outcomes – national and local datasets, part 3. MHCLG, London, UK. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/694356/20180326_Troubled_Families_-_families_outcomes_report.pdf

Moser, C., Kalton, G., 1971. Survey methods in social investigation. London: Routledge.

National Housing and Planning Advice Unit (NHPAU), 2009. A guide to the Reading-CLG Affordability model. Office for National Statistics, London UK. Available at: https://www.google.co.uk/search?hl=en&q=Segensworth+Road%2C+Titchfield+Fareham+PO15+5RR&meta=

Waugh, A., Clark, A., Knowles, J., Rowley, D. 2018. Health and homelessness in Scotland.People, Communities and Places. Scottish Government. Available at:

https://beta.gov.scot/binaries/content/documents/govscot/publications/research-publication/2018/06/health-homelessness-scotland/documents/00536908-pdf/00536908-pdf/govscot:document/

Yates, F., 1981. Sampling methods for censuses and surveys. Charles Griffin and Company Limited.

Appendix: Data overview

A.1 Existing data sources

A.1.1 Data on homelessness and rough sleeping

Data on homelessness are mainly collected from local authority (LA) registries and cover statutory homeless populations. Current homelessness estimates based on LA administrative data only include homeless households considered to be unintentionally homeless and that fall in a priority need group (e.g. having dependent children or being vulnerable because of mental health illness). Under the new Homelessness Reduction Act 2017, LA homelessness duties have been expanded to cover all eligible households regardless of whether they are intentionally or in priority need. A new strategy for data collection (H-CLIC) has been adopted to reflect this change in policy.

Data on non-statutory homeless groups (particularly rough sleeping populations) are mainly collected from LAs, agencies and charities that provide emergency and temporary accommodation as well as prevention services. For example, national statistics on rough sleeping and statutory homelessness (based on the official definition of statutory homelessness prior to the Homelessness Reduction Act 2017) are derived from rough sleeping counts and estimates as well as numbers of people in temporary accommodation that LAs report to MHCLG. In addition, MHCLG administers a questionnaire about rough sleeping populations to all homelessness services providers that are funded through the Homelessness Prevention Programme. These data, which are primarily assembled to assess the effectiveness of these services, can be used to determine the size of rough sleeping populations.

Non-statutory homeless populations can also be observed in temporary accommodation and hostel administration data. For example, the Homeless Link charity has created a database that compiles information on rough sleeping referrals, bedding capacity and groups that have access to accommodation services from approximately 1,400 accommodation projects and day centres. Although the data is not live, it is updated regularly and is the most accurate source of data about homelessness services in England.[31]

[31] See here for more information http://www.homeless.org.uk/search-homelessness-services

Administrative data on homelessness (P1E data and Homelessness Case Level Information H-CLIC)

P1E data

Under the previous homelessness and housing acts (for example, the Homelessness Act 2002 and Housing Act 1996), LAs were responsible for providing homelessness support according to a set of eligibility criteria, which focused on households that are unintentionally homeless and fall in a priority need group (families with dependent children or pregnant women and people who are vulnerable in some way, e.g. because of mental illness, time previously spent in care, custody or the armed forces, or having to flee their home because of violence or the threat of violence). All statutory homeless households that have applied for homelessness duties and fulfil these criteria (*acceptances*) are offered suitable, long-term accommodation. Where settled accommodation is not available immediately, the LA will place the household into Temporary Accommodation until it becomes available. Applicants that are considered to be intentionally homeless – for instance, because they have a house that they choose not to live in – or not in priority need are offered assistance to find accommodation on their own.

Prior to the introduction of the H-CLIC collection system, data on statutory homeless households that were offered temporary accommodation were collected using the PE1 form on the last day of each quarter.[32] Information was collected on

- *number of applications and decisions* (for example, acceptances, in priority need but intentionally homeless and unintentionally homeless but not in priority need)
- *demographic characteristics of acceptances* (gender, age, ethnicity, household type, number of children),
- *priority need category* (e.g. left home because of an emergency, family with dependent children or pregnant women, applicant aged 16 or 17 years old, applicant between 18-20 years old who was previously in care, being drug and/or alcohol dependent, vulnerable because has been previously in care, custody or armed forces, had to flee home because of violence)
- *main reason for loss of previous house* (overstaying one's welcome, violence, harassment and intimidation, rent arrears, termination of tenancy, exit from institution or other LA care),
- *immediate outcome of application* (including temporary accommodation and provision of other services for homelessness prevention and relief – e.g. support to find accommodation in the private rented sector),
- *number of homelessness duties*

[32] LA data on homelessness under the previous acts can be found here: https://www.gov.uk/government/statistical-data-sets/live-tables-on-homelessness#homelessness-summary-local-authority-level-tables

that have ended during the quarter (and reason for ending)

- *number of applications and decisions for returning UK national and foreign national applicants*

The P1E forms also collected information regarding homelessness *prevention* (providing people with support and advice services to avoid the risk of homelessness) and *relief* (when the authority is unable to prevent homelessness but helps someone to secure accommodation even if they do not fulfil the eligibility criteria) achieved by LA services. The following three types of successful cases were recorded in the P1E form:

- cases where homelessness is prevented and applicants remain in their home (and type of prevention service – e.g. financial payments from homelessness prevention fund, debt advice, resolving housing benefits problems),
- cases where homelessness is prevented through the provision of assistance in securing accommodation (and type of service – e.g. hostel, private rented accommodation, social housing, low-cost home ownership scheme), and
- non-priority or intentionally homeless cases that had their homelessness relieved through receiving assistance to secure accommodation (and type of service – e.g. hostel, private rented accommodation, social housing, low-cost home ownership scheme).

The last statutory homelessness statistics based on the P1E returns were published in June 2018.

H-CLIC data

New LA duties that will not be limited to priority cases were introduced in the Homelessness Reduction Act 2017. In the context of the new Homelessness Reduction Act (April 2018), LAs will provide prevention and support (*'relief'*) services to all those eligible for public services even if they do not fall in the priority categories. The HRA extends the period over which households can be owed a prevention or relief duty from 28 to 56 days. Therefore, P1E data do not cover the new legislation for homelessness duties offered by the LAs.

This new legislation led to the collection of new case level homelessness data (H-CLIC) that is expected to contribute to better understanding of what causes various types of homelessness, what are the effects of homelessness on the individual and household level and what are the best prevention and support measures.

The H-CLIC data collection system is introduced to cover applicants – for the extended period of 56 days - to LA services that fall in the statutory homelessness category irrespective of whether they comply with the *priority needs* and *unintentionally homeless* definitions. If support is needed beyond 56 days, LAs will apply *priority needs* and *unintentionally homeless* definitions to determine whether a statutory duty is still owed. H-CLIC data are collected at the end of each quarter and include dates of activities reported during the quarter.

The database is expected to cover a set of factors that are shown to predict homelessness – such as other support needs and unemployment – as well as effective ways to prevent and relieve homelessness. Specifically, the H-CLIC system goes beyond the P1E form by collecting new data in the following areas:

- socio-demographic characteristics of applicant (sexual orientation, employment status),
- socio-demographic characteristics of all household members (age, gender, relationship with applicant, employment status where applicable),
- welfare benefits (benefits aiming to support housing costs and other living costs),
- assessment of type of priority need (for example, if the household includes dependent children or applicant is/household includes a pregnant woman, etc.),
- type of last settled accommodation (long-term accommodation, assured short-hold tenancy, supported housing),
- support needs for main applicant and other household members (for example, mental illness, drug and alcohol abuse or being vulnerable because of time spent in armed forces) and assistance provided by LAs,
- date(s) of entry (exit) into (out of) temporary accommodation,
- length of time of prevention and relief services and whether they were successful or not,
- whether or not a case was subject to review and if the review was successful.

Finally, each household that has applied to LAs for homelessness is assigned a unique identifier. This indicator can be used to accommodate future linkage of H-CLIC data to observations from other administrative sources (for instance, benefits, health, education and child safeguarding).

MHCLG plans to collect personal data from homelessness services applicants in the context of the H-CLIC to observe homelessness experiences over time. However, whether the underlying projects will obtain the legal sign-off required for collecting personal information and thus adding a longitudinal element to H-CLIC is yet uncertain.

LA counts of rough sleeping populations

Official statistics on rough sleeping populations are collected by LAs and reported to MHCLG on an annual basis. LAs either count or estimate (potentially including a spotlight count when necessary)[33] the number of people who sleep rough at a single night in their area. When street counts are not conducted, LAs collaborate with all the organisations that support rough sleepers in the area (even if not directly targeting rough sleeping groups – for example, mental health services) to arrive at a reliable estimate of people who sleep rough at a single night in the LA area.

Figures on rough sleeping represent single snapshots without

[33] See the Homelessness Link guidance for more information on collecting data on rough sleeping populations across LAs: https://www.homeless.org.uk/sites/default/files/site-attachments/Counts%20%26%20Estimates%20Introduction%202018.pdf

distinguishing between rough sleeping stocks, flows and returns. Information is collected for a set of demographic characteristics of rough sleeping groups – particularly, age, gender and nationality.

Every organisation that provides services to rough sleeping populations participates in Homeless Link, a national charity which validates and checks the accuracy of the rough sleeping counts provided by LAs. Moreover, Homeless Link guides the LAs through the process of counting or estimating rough sleeping populations. Independent verification and data collection guidelines that are common across LAs improve the consistency of rough sleeping statistics across areas.

However, it appears that official counts of rough sleeping groups by LAs tend to underestimate the population of people who sleep rough. For example, the official statistic on the size of the English rough sleeping population in 2016 (4,134 people sleeping rough on a single night between 1 October and 31 November)[34] is lower than the estimated level of rough sleeping (around 8,000) for the same year from research conducted by Bramley (2017) on behalf of Crisis. It is likely that official counts underestimate the true size of rough sleeping due to potentially ineffective data collection methods as well as limited collaboration between LAs and other public services that offer support to people who are sleeping rough. However, evidence suggests that despite their limitations, official counts are indicative of existing trends in rough sleeping.[35]

Further data collections on rough sleeping populations

Combined Homelessness and Information Network (CHAIN)

The Combined Homelessness and Information Network (CHAIN) is a multi-agency comprehensive database funded by the Greater London Authority (GLA) that assembles information about people who sleep rough in London.

CHAIN compiles data from various organisations that provide support to rough sleeping groups in London. Local Authorities, outreach teams, accommodation programmes, day centres and assessment and reconnection interventions such as the No Second Night Out (NSNO)[36] programme share data regarding the work they do with people who rough sleep as well as their demographic characteristics and needs.

Statistics are reported bi-monthly, quarterly and annually across boroughs in London and are used to evaluate existing services and identify

[34] See here for more information on population of people sleeping rough in 2016: https://www.homeless.org.uk/sites/default/files/site-attachments/Homeless%20Link%20%20analysis%20of%20rough%20sleeping%20statistics%20for%20England%202016.pdf

[35] See here for more information: https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-assessmentreport320statisticsonhomelessnessandroughsleepinginenglan_tcm97-45078.pdf

[36] See here for more information regarding the No Second Night Out programme: http://www.nosecondnightout.org.uk/

rough sleeping trends and needs.

CHAIN records information on people who fall into three categories: i. new rough sleepers (those identified as sleeping rough but who have not yet been contacted by outreach teams), ii. living on the streets (those having high number of contacts during the last three weeks which suggests that they live on the streets), and iii. intermittent rough sleepers (people who have been contacted during the period of data collection but not enough to be considered as living on the streets). However, the data do not cover hidden homeless groups that cannot be reached by outreach teams – for example, those living in squats or places that are not known or accessible by outreach workers.

CHAIN data covers the following areas:

- *counts of rough sleeping populations* (number of people seen rough sleeping – flows, stocks, returners, number of times seen rough sleeping)
- *people seen rough sleeping for the first time* (counts, accommodation prior to rough sleeping, institutional history, reasons for leaving last accommodation – eviction, loss of job, financial problems, relationship breakdown, violence, end of stay in short/medium term accommodation or institution, housing conditions),
- *demographic characteristics* (nationality, gender, age, ethnicity),
- *support needs* (mental health problems, drug and alcohol abuse),[37]
- *institutional history* (armed forces, prison, social care),
- *accommodation* (booked into long-term or temporary accommodation) and *reconnection* (return to home area, seeking work, move to area where family and friends are, move to area with appropriate services) outcomes, and
- *temporary accommodation* outcomes (arrivals and departures – destination of departure and reasons for leaving).

Homeless Link and Street Link databases

Homeless Link – a charity representing organisations across the country that support and provide accommodation for people sleeping rough – has a dataset that includes information about accommodation and non-accommodation programmes for people who sleep rough.

Accommodation projects primarily offer hostel accommodation to people who sleep rough. Information is collected on hostel location (for example, addresses and LAs), number of beds, disadvantaged groups that have access to the hostel (including homeless and rough sleeping, among other groups) and minimum and

[37] More detailed information such as support needs are only recorded for those rough sleepers who are in regular contact with outreach workers (for example, many people might sleep rough only one or two days while others are sleeping rough more regularly).

maximum length of time that people can stay at the hostel.

Homeless Link shares StreetLink referrals data with MHCLG, which includes monthly number of referrals, types of referrals, and referral outcomes as well as some socio-economic characteristics of the people who sleep rough – including age and location (LA, region, postcode etc.)

StreetLink is a website, mobile app and phone service that enables members of the public to alert local authorities and street outreach services in England and Wales about people they have seen sleeping rough. It is run in partnership by Homeless Link and St Mungo's and is funded by grants from MHCLG, the Greater London Authority and the Welsh Government.

A.1.2 Costs of homelessness and housing services

Information about the costs of homelessness and housing services is central to the assessment of policies that aim to tackle homelessness and provide support to those who are homeless.

The Unit Cost Database[38] comprises estimated social and economic costs in various areas including housing and social services. The following average values of costs related to homelessness are documented:

- eviction (average fiscal cost of complex evictions, single repossession),
- homelessness applications (on-off and ongoing costs),
- temporary accommodation (cost of housing a homeless household in hostel accommodation),
- advice and prevention support (per scheme),
- rough sleeping (LA expenditure per individual), and
- housing benefits (expenditure on benefits and cost of processing an application).

The database can be used to forecast costs and benefits associated with housing interventions and projects and to perform Cost-Benefit Analysis (CBA). Data collections from homelessness services (prevention and temporary accommodation services) can be measured against reported costs to assess the costs of providing support to homeless populations as well as potential cost reductions from successful prevention.

Homelessness and rough sleeping are expected to entail wider costs that go beyond funds spent on homelessness prevention and treatment services per se. A review of the evidence on costs of homelessness conducted by MHCLG[39] suggests that homelessness and rough sleeping costs are

[38] For more information about the database see here: http://www.neweconomymanchester.com/our-work/research-evaluation-cost-benefit-analysis/cost-benefit-analysis/unit-cost-database

[39] For the full report on homelessness costs see here: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/7596/2200485.pdf

composite and are likely to include expenses on welfare benefits, use of health care services and substance use treatment programmes as well as costs associated with policing services and the social justice system.

A.1.3 Survey data

Longitudinal survey data can be used to observe patterns of homelessness over time and identify the set of factors that have the largest effects on homelessness. Therefore, they are useful for producing estimates of future levels of homelessness and rough sleeping conditional on sets of determinants as well as gauge homelessness levels for specific population segments. Survey data can feed into behavioural models included at the first stage of simulation models to quantify links between sets of determinants and outcomes of interest. Moreover, micro-simulation models can be applied to data at the individual level drawn from surveys to produce projections of homelessness and rough sleeping that can be broken down by specific socio-economic characteristics (e.g. gender).

Survey data can also be used as a source of additional detailed information to complement analysis based on administrative data. For example, being able to link individuals' history of service use and welfare benefits uptake using registry data to their personal circumstances – such as mental health and relationship with family and friends – using survey data would allow for a comprehensive assessment of pathways in and out of homelessness and rough sleeping. Moreover, survey data can be used to triangulate results on types of homelessness which are likely to be underestimated in official counts. For example, while count data on people who live in hostels and other types of emergency accommodation is available, it is likely to be limited and non-representative. In this case, we can identify hostel residents using survey data and arrive at representative counts based on survey weights. Moreover, the length of stay at hostels and other types of emergency accommodation, which is an important factor when measuring homelessness, can be estimated using survey data.

Surveys are primarily developed for the purpose of statistical analysis – therefore, their design addresses issues such as population coverage and representativeness as well as consistency of reported information (Moser and Kalton, 1971; Yates, 1981). For example, there is a set of sampling and weighting techniques that can be applied to make sure all subpopulations of interest (e.g. age groups, disadvantaged groups, women, etc.) are covered and the survey sample is representative of these populations. Overall, while the risk of bias (e.g. self-response bias, selection bias and memory bias) cannot be fully eliminated, using appropriate methods to analyse survey data usually results in reliable estimation of the links of interest (Kelly et al., 2003).

A major limitation of surveys is that population subgroups that experience or are at risk of experiencing homelessness are essentially hard to capture and might thus be under-represented in the survey samples. For example, existing surveys usually omit groups of sofa surfers as they do not collect information on people who are temporarily staying in private households. It is not an easy task to contact and survey homeless populations as they are frequently socially excluded, vulnerable and transient. Therefore, the coverage of such populations in cohort and household surveys is likely to be limited.

Moreover, attrition, which is likely to be higher among disadvantaged groups, might result in people in homelessness or at risk of homelessness dropping off longitudinal surveys such as the Understanding Society survey. In this case, cross-sectional surveys such as the English Housing Survey, which collect information on representative samples that are different in each wave, are more appropriate for counting homeless populations.

Household surveys, such as the English Housing Survey (EHS), do not capture people who are currently homeless – even if a member of a surveyed household is currently homeless, it is unlikely that enumeration teams will be aware of this or able to contact them. Such surveys are mainly limited to recording only *past* experiences of homelessness and rough sleeping. For example, the new module on homelessness experiences, which was added to the EHS in the 2016/17 wave, asks respondents whether *'they have ever contacted the council because they were homeless or about to be homeless in the last years*'.[40] Moreover, the Survey of Living Conditions (as part of the EU-SILC) will include a module on past experiences on homelessness in 2018.

Specialist surveys that are developed to collect data from these populations present a way to address this limitation. For example, the Multiple Exclusion Homelessness Survey (MEHS) explicitly targets groups of people who use homelessness services such as day centres and emergency accommodation. Another example is the Poverty and Social Exclusion Survey (PES) that aims to gather evidence on a set of life domains from vulnerable groups who live in poverty and deprivation. PES draws on the sample of households that have participated in the Family Resources Survey (FRS) to select a subsample of households that fall in the bottom income quintiles – FRS serves as a sampling frame to oversample disadvantaged households of interest (Gordon, 2011).

Another important point to consider is the level of data disaggregation in surveys. The ability to identify groups of observations across LAs in England is essential for applying models that predict homelessness experiences conditional on housing and labour market influences which operate at the local level (e.g. Housing Market Area – HMA or Local Labour Market Area – LLA, which can be approximated by single LAs or groups of LAs). Moreover, disaggregated data are important for assessing the

[40] See here for the EHS questionnaire documentation 2016-17: http://doc.ukdataservice.ac.uk/doc/8384/mrdoc/pdf/8384_ehs_questionnaire_documentation_year_9_2016_17.pdf

effectiveness of interventions that are mainly implemented at the LA level. While household and cohort surveys usually record information on the location of study subjects (e.g. postcode, LA, region), it is not always straightforward to get access to geography identifiers. A special license is needed to access geographical indicators in the majority of surveys while in some cases (such as the EHS), this information is not attached to the datasets. Moreover, in some cases, the size of the LA-specific sample might be too small to allow for such a comprehensive analysis.

Here, we describe a list of available surveys that capture experiences of homelessness and rough sleeping among other life domains.

Multiple Exclusion Homelessness Survey (MEHS)

The Multiple Exclusion Homelessness Survey (MEHS) is a small-scale study of vulnerable groups that are likely to experience deep social exclusion. The survey particularly targets people with experiences of multiple exclusion homelessness – defined as a form of social exclusion involving homelessness and one or more of the following situations: drug and alcohol abuse, history in institutional care and ‘street living activities’ such as begging and street drinking (Fitzpatrick et al., 2012).

Data were collected from all users of randomly selected ‘low threshold’ services[41] aiming to provide support to people who face various aspects of exclusion such as homelessness and drug abuse. The survey was conducted in seven cities in the UK (Belfast, Birmingham, Bristol, Cardiff, Leeds and Westminster) over two weeks.

At the first stage of the survey (Census Questionnaire Survey), questionnaires were assigned to all users of ‘low threshold services’ – data were collected by more than 1,000 users who returned the questionnaires. Service users were mainly asked questions regarding their demographic characteristics and use of ‘low threshold’ services for a set of reasons including homelessness, drug treatment.

At the second and main stage (Extended Interview Service), approximately 450 study subjects from the Census Questionnaire Survey who were found to experience episodes of multiple exclusion homelessness were further asked a set of questions that cover the following areas:

- *homelessness* (stayed at a hostel, foyer, refuge, night shelter or B&B hotel; stayed with friends or relatives – sofa surfed; slept rough; applied to the council as homeless),
- *drug and alcohol dependence and abuse* (period when had 6 or more alcoholic drinks per day; used hard drugs; injected drugs, etc.),
- *institutional history* (went to prison; admitted to hospital because of mental illness; left LA care),
- *street culture activities* (shoplifting;

[41] Low threshold’ services are those that provide immediate support to vulnerable groups without aiming to offer long-term solutions. For example, ‘low threshold’ drug treatment programmes are interventions that provide support and counselling to people with drug dependence problems without demanding that users control their drug intake. Day centres, soup runs, emergency accommodation, street outreach teams and drop in services are ‘low threshold’ homelessness services.

street drinking; begging, etc.),

- *adverse life events* (divorced; evicted of rented property; thrown out by parents; redundancy; bankruptcy, etc.),
- *extreme exclusion/distress* (period where very anxious and distressed; victim of violent crime; sexual assault as a child, etc.), and
- *adverse circumstances in childhood* (run away from home; didn't get along with parents, homeless family; neglect, etc.).

While it does not cover the full set of potential predictors of homelessness (for example, poverty) and only includes a sample of small size, the second part of the MEHS includes information that is central to understanding why homelessness occurs for vulnerable and disadvantaged groups and identifying pathways into homelessness for groups with complex needs. The questionnaire could potentially be expanded to include more information on homelessness determinants and use of public services around homelessness and rough sleeping. There is merit in using an extended questionnaire in a future MEHS wave to collect information from larger samples of services users. Moreover, a future MEHS wave could be conducted jointly or linked with other surveys such as the UK Destitution Survey, which adopts the same methodology for data collection to survey larger target groups.

UK Poverty and Social Exclusion Survey (PSE)

The Poverty and Social Exclusion Survey (PSE) aims to improve the measurement and understand the extent and nature of poverty, inequality, deprivation and social exclusion as well as the influence of welfare policies implemented to support disadvantaged groups in the UK (Gordon, 2016). It consists of a quantitative part where data are collected from samples of disadvantaged households and individuals and a qualitative part that aims to add more depth to our knowledge about experiences of poverty.

The survey was introduced in 1979 – three additional waves were conducted in 1983, 1990 and 1999. Recently, two waves have taken place in 2010 and 2012 drawing on samples of respondents from the 2010/11 Family Resources Survey (2012) who have agreed to be contacted again). The PSE sample was selected based on known characteristics about income, area of residence and demographic characteristics such as ethnicity. An equivalised income variable was created using the raw FRS data to sample respondents from the entire range of income groups (Maher and Drever, 2013).

PSE includes identifiers for experiences of various types of homelessness among other experiences of disadvantage. Specifically, PSE respondents were asked whether they have ever been homeless and had to stay in family and friends' place, insecure or temporary/emergency accommodation or sleep rough during the last five years or more than five years ago.

The questionnaire also asks respondents about characteristics and circumstances that are shown to predict homelessness. The following topics are covered: household composition and changes, nationality, ethnicity, poverty and deprivation, use of local public services, financial problems and debt, employment conditions, health, disability, social networks and support, crime and criminalisation, and critical life events such as left parental home, got divorced or separated, lost or left job, had an important health problem, etc.

British Cohort Study (BCS)

The 1970 British Cohort Study (BCS70) is a longitudinal survey that interviews at regular time intervals a sample of individuals who were born during a single week in 1970 in England, Scotland, Wales and Northern Ireland.

In the 2000 BCS wave, a questionnaire was introduced in the survey that included information about experiences of homelessness. Homelessness is defined as *"having moved out of a place and having nowhere permanent to live"*. This definition is quite broad, covering various types of homelessness (e.g. rough sleeping, sofa surfing, living in hostels, etc.) – however, there is no way to distinguish between different types of homelessness. In particular, the questionnaire included the following questions:

- whether study subjects had experienced any homelessness incidence between 1986 and 1991,
- when did the event take place,
- the main reasons they had to move out of the accommodation before becoming homeless,
- where did they stay while homeless, and
- how long they have been homeless for.

The questionnaire includes questions that cover a set of characteristics and circumstances that may predict homelessness – for example, past and current relationships, children, family relationships and support, income, employment and health.

The following waves (2004, 2008, 2012 and 2016) include questionnaires that ask participants if they are currently homeless or have been homeless since the previous wave and record the length of the homelessness spell. The survey does not collect information on single homelessness (e.g. sofa surfing, staying in hostels, etc.) However, they do not include the level of detail integrated into the 2000 questionnaire. Therefore, the findings cannot be compared between waves.

Millennium Cohort Study (MCS)

The Millennium Cohort Study (MCS) is a multi-disciplinary research project following the lives of around 19,000 children born in the UK in 2000-01. It is the most recent of Britain's national longitudinal birth cohort studies. Study subjects and their families are regularly interviewed (for example, in 2001-2, 2004-5, 2006, 2007, 2008, 2012 and 2015). The last wave was conducted in 2018 – the data from the latest wave will be available at the end of 2019.

MCS covers a broad set of topics, including family context, education and schooling, parenting activities, parents' employment and income, health, housing conditions and characteristics of their area of residence.

The parents are also asked whether they have experienced any incidence of homelessness since the last time they were interviewed. MCS reports experiences of family homelessness rather than single homelessness – the parents of the study subjects are asked if they had to leave their house and had nowhere permanent to go. As in the BCS questionnaire, there is no way to distinguish between different types of homelessness.

Information is reported regarding the timing of the homelessness spell, the main reasons they had to leave the accommodation before becoming homeless, the places they stayed while homeless and the length of the homelessness spells. The same questions are included in each wave, allowing for longitudinal observations of paths in and out of homelessness and estimation of links between explanatory variables and homelessness occurring over time.

English Housing Survey (EHS)

The English Housing Survey (EHS) is a national survey commissioned by MHCLG that takes place annually. It collects information about several topics related to demographic characteristics and personal circumstances (e.g. age, gender, nationality, ethnicity, religion, disability, education and health), employment and earnings, pensions, benefits and income support, savings and investment as well as housing circumstances and conditions.

A new set of questions about homelessness were added to the survey in the year 2016/2017. The latest questionnaire asks study subjects the following questions:

- whether they ever contacted LAs because they were about to become homeless,
- whether they asked the council to consider them as homeless,
- whether they were accepted by the council (considered in priority need, offered accommodation),
- what type of accommodation did the council offered them (emergency housing, temporary housing, council accommodation, long-term housing association accommodation), and
- whether the council offered any help or advice (financial, referral to a secured shorthold tenancy, advice with rent, housing benefits, issues with landlord).

This set of questions about homelessness was added only in the latest wave, which does not allow for longitudinal analysis of homelessness experiences. The findings from this new module will be available in July.

The survey includes additional questions related to housing issues such as:

- rent and housing benefits,
- current tenancy agreement,
- past tenancy and deposit,
- type and size of accommodation,
- housing history, council tax and utilities, and
- satisfaction with accommodation.

Understanding Society – the UK Household Longitudinal Study (UKHLS)

The Understanding Society (UKHLS) survey is a revised and updated version of the British Household Panel Survey (BHPS), created in 2009 to collect data from the same large and nationally representative sample of households across the UK on an annual basis. The UKHLS datasets include a harmonised version of the BHPS waves for the years prior to the introduction of US (1991-2009). The UKHLS surveys all household members that are aged ten years or older from around 40,000 households in the UK.

The UKHLS does not include a straightforward identifier of homeless households or members of households. However, study subjects who are provided accommodation by LAs or housing associations and charities can be identified in the sample. Eviction episodes can also be tracked over time.

The dataset also covers an array of topics that are related to homelessness, including household composition, housing conditions, residential mobility, education, health and usage of public services, labour market outcomes and income from employment, pensions and welfare benefits including housing-related benefits. There are also retrospective questions related to lifetime history of personal relationships (such as marriage and family), labour market trajectories, wealth and assets, and health and ageing.

Finally, the UKHLS is the largest household-based survey providing crucial information about important issues (for example, life attitudes, finance, health, employment) for the whole population in the UK. Therefore, there is scope to include direct questions about homelessness utilising the longitudinal nature of the survey to better understand how people reach the state of homelessness, what factors trigger homelessness episodes and what population subgroups are more at risk.

Joseph Rowntree Foundation (JRF) Destitution UK Survey

The Destitution survey aims to understand material and income deprivation in the UK and to explore experiences of destitution among vulnerable groups. The definition of destitution has two elements: the first is based on specific material deprivation – lacking two or more of the following six essentials over the past month because individuals cannot afford them: shelter, food, heating home, lighting home, clothing and footwear, and basic toiletries (Fitzpatrick et al., 2018). The second is based on low income and absence of savings.

This survey, commissioned by the Joseph Rowntree Foundation (JRF), collected data from users of 16 crisis services across the UK in 2015 and 2017. The services that are included in the sample were selected to ensure that satisfactory ranges of destitution experiences, migrant populations and urban/rural shares are covered. In 2017, the sample included approximately 3,000 unique service users.

The survey can potentially be used to

observe rough sleeping groups with complex needs. National estimates of people who experience destitution are estimated based on survey data. In addition to counts of service, the questionnaires cover the following areas:

- income of destitute households,
- household type (single-working, single older, couple, lone parent, couple family, multi-adult),
- experiences of severe poverty,
- migration profile,
- current living arrangements (private housing, rough sleeping),
- housing tenure, and
- sources of financial support.

Moreover, survey data and official statistics are analysed jointly to generate indicators for risks of facing destitution with respect to specific predictive factors.

Rough sleeping estimates using data from the JRF Destitution Survey are substantially higher than the official counts. On the other hand, the estimated size of the population in hostels and other types of emergency accommodation triangulates reasonably well with estimates from the Homeless Link. One limitation of the JRF Destitution Survey is that it does not cover groups that are in hidden homelessness (e.g. sofa surfers) – therefore, there is merit in revising and updating the survey's design and data collection tools (e.g. questionnaires, sampling techniques, etc.) in order to cover these groups.

A.1.4 Data on related factors

Simulation models take into consideration the influence of a set of economic, personal and policy-related factors to arrive at longer-term estimates of future trends in homelessness under neutral (baseline – assuming no changes in policy) or planned policies scenarios. The models include a set of covariates for two purposes: i. to estimate behavioural responses to changing covariates and quantify links between covariates and outcomes of interest at the individual or local level (first stage), and ii. to project homelessness in the future conditional on expected trends in explanatory variables based on estimated elasticities either from the first stage or empirical literature (simulation stage).

Table 1 presents a set of covariates that might be used to predict future trends in homelessness for services delivery planning or appraisal purposes. The covariates are classified into categories, and a list of relevant variables and potential data sources is shown for every category. While the list should not be thought of as exhaustive, it is suggestive of what sources of data are available and can be used as inputs to a suite of models around homelessness.

As discussed in the previous section, data at the individual or household level – for instance, the data from the Labour Force Survey (LFS) and the DWP data on benefits – are useful to quantify behavioural responses to changes in demographic, economic, personal and policy factors, such as reduction in welfare benefits, household poverty, or experiences of relationship breakdown.

However, to estimate longitudinal links between explanatory factors and outcomes of interest, surveys that cover a wide set of areas including homelessness and other important covariates should be used – for example, BCS and MCS. This might restrict the estimation of links of interest as household surveys often face a number of limitations regarding the coverage of homeless populations.

A potential way to address this limitation is to use individual observations from different sources that are linked with unique person identifiers when such linkages are available. For example, datasets from government sources that include information on homelessness such as the National Drug Treatment System and Public Health Outcomes Framework can be linked to data on homelessness provided by the LAs.[42]

Additionally, the models can use a series of national statistics on key indicators – including house prices and labour market outcomes – to estimate links between homelessness levels and explanatory variables. It should be noted here that disaggregation of outcomes at the LA level is central to models around homelessness used to inform planning decisions for delivery of treatment and prevention interventions.

Therefore, data on variables that are relevant to LAs such as affordability ratios, employment, housing supply (measured by counts of new residential dwellings) and health indicators should be inserted into the models. National statistics at the LA level can feed into either a complex model generating outputs at the national level that can be further disaggregated at the subnational level, or versions of a simpler model that can be estimated separately for each LA.

[42] We further discuss data linking in the recommendations section (2.3.1). Examples of data linking processes are presented in tables 4 and 5.

Table 2. Data sources for relevant factors

CATEGORIES OF RELATED FACTORS	VARIABLES	SOURCES	REPORTING FREQUENCY	DATA COLLECTION UNIT	GEOGRAPHICAL GRANULARITY
HOUSING MARKET	House price indices	HM Land Registry data	Monthly	Residential property (census – when sales transaction takes place)	Sub-regional (LAs)
	House prices	ONS House Price Statistics for Small Areas (HPSSAs)	Annual (updated quarterly)	Residential property (census – when sales transaction takes place)	Sub-regional (LAs)
	Private rents	Valuation Office Agency – Private rental market statistics	Monthly	Private residential property (sample of rental information)	Sub-regional (LAs)
	Housing affordability (ratio of housing prices to annual earnings)	ONS Housing Affordability statistics (HPSSAs & ASHE – annual snapshot of earnings)	Annually	Complex statistics	Sub-regional (LAs)
	LA social housing (dwelling stocks, Right to Buy applications & sales, Social Homebuy sales, social housing lettings, vacant dwellings, condition of dwelling stocks, rents, rent arrears, affordable housing)	MHCLG Local Authority Housing Data (LAHS)	Annually	LA owned and managed dwelling (census)	Sub-regional (LAs)
	New build dwellings	MHCLG House Building data	Quarterly	Dwelling (census)	Sub-regional (LAs)
	New planning applications	MHCLG Planning Application statistics	Quarterly	Residential planning application (census)	Sub-regional (Local Planning Authorities – LPAs)
	Social housing lettings (characteristics of households – age, ethnicity, economic status, previous tenure – and properties, rents)	MHCLG Continuous Recording (CORE) system	Annual	Household/property (census – each time property is let)	Sub-regional (LAs)

CATEGORIES OF RELATED FACTORS	VARIABLES	SOURCES	REPORTING FREQUENCY	DATA COLLECTION UNIT	GEOGRAPHICAL GRANULARITY
LABOUR MARKET	Employment, employment types (part-time, temporary employment), hours worked, unemployment, labour market participation, redundancies, occupation, earnings	Labour Force Survey (LFS)	Quarterly	Household/ individual (sample)	Regional (open access) – LA level (special license)
	Earnings (distribution), occupation, hours worked	Annual Survey of Hours and Earnings (ASHE)	Annually	Individual (sample of jobs drawn from HR Revenue & Customs and PAYE records)	Sub-regional (LAs)
	Employment, types of employment	Business Register and Employment Survey (BRES)	Annually	Business (sample)	Sub-regional (LAs)
BASIC ECONOMIC INDICATORS	GDP	ONS Statistics	Monthly & Quarterly	Complex estimate	National
	GVA	ONS Statistics	Monthly	Complex estimate	Sub-regional (LAs)
	Consumer price inflation (CPI), Consumer Prices Index including owner occupiers' housing costs (CPIH)	ONS Consumer Price Indices	Monthly & quarterly	Complex estimate	Regional
	Public finances (borrowing, deficit, debt)	ONS Public Sector Finances	Monthly	Central government, local government (census)	Local government
	Household debt (including residential mortgages) – value of loans to individuals	Bank of England – Lending to individual data	Monthly & quarterly	Complex estimate (bank data)	National (disaggregation by type of lending – e.g. lending secured on dwellings)
POVERTY	Household incomes, income inequality, poverty, child poverty, in-work poverty	IFS Living Standards, Inequality and Poverty Spreadsheet (data from DWP series, Family	Annual	Complex estimate	Regional (three-year averages)

CATEGORIES OF RELATED FACTORS	VARIABLES	SOURCES	REPORTING FREQUENCY	DATA COLLECTION UNIT	GEOGRAPHICAL GRANULARITY
		Resources Survey data and Understanding Society)			
WELFARE BENEFITS	Housing benefit by family type, tenure, and award	DWP Housing Benefit caseload statistics	Monthly	Household/individual claimant	Sub-regional (LAs)
	Job seekers allowance	DWP Job Seekers Allowance data (JSA)	Quarterly	Individual claimant	Sub-regional (LAs)
	Universal Credit (age, employment status, conditionality regime)	DWP Universal Credit statistics – will replace other benefit datasets in 2019	Monthly & quarterly	Individual/Household claimant	Sub-regional (LAs)
	Personal independence payment, disability living allowance (disability condition, age, gender)	DWP Personal Independence Payment Statistics	Quarterly	Individual	Sub-regional (LAs)
MENTAL AND PHYSICAL HEALTH	NHS secondary mental health services use, mental health assessment (number of people in NHS mental health services, average length of stay)	Mental Health Minimum Dataset (MHMDS)	Annual (2002-2008)	Individual/secondary health services users	National
	Mental health, learning disability, autism services (NHS services)	Mental Health Services Dataset (MHSDS)	Monthly	Individual (services user)	Sub-regional (LAs)
	Numbers of users in drug treatment programmes (and outflows)	Home Office National Drug Treatment Monitoring System	Monthly	Individual (service user)	Sub-regional (LAs)
	Health status of rough sleepers, cost of health care services, accommodation type	Hospital Episodes Statistics (HES)	Monthly	Individual (service user)	National *(geographical information such as where patients are treated, and the area where they live is included)*

CATEGORIES OF RELATED FACTORS	VARIABLES	SOURCES	REPORTING FREQUENCY	DATA COLLECTION UNIT	GEOGRAPHICAL GRANULARITY
	Health inequality, health protection, healthcare	Public Health England (PHE) Public Health Outcomes Framework	Quarterly	Complex estimate	Sub-regional (LAs)
	Health indicators (suicide rate, mortality rate, life expectancy, alcohol-specific hospital stays, excess weight, injuries and ill health, child health, health inequalities,)	PHE Local Authorities Health Profiles (LAHP)	Annual	Complex estimate	Sub-regional (LAs)
VULNERABLE GROUPS	Care leavers (aged 17,18,19,20,21): type of accommodation, suitability of accommodation	Department for Education - Children Looked After in England	Annual	Individual (child in care)	Sub-regional (LAs)
	Prisoners: accommodation status on entry to custody, release from custody, start and end of community orders.	Accommodation Status of Offenders (collected by probation providers)	*Experimental data collection (not yet a dataset)*	Individual (prisoner)	-
DEMOGRAPHICS	Population, migration (age and gender breakdown)	ONS Population Estimates	Annual	Complex estimate	Sub-regional (LAs)
	Household characteristics (age-type, tenure status), household reconstitution and ageing	Labour Force Survey (LFS)	Quarterly	Household/individual	Regional (open access), LA (special license)

Box A1. Forecasts of related factors

Projecting homelessness trends in the medium to longer term and evaluating potential effects of changes in policy and economic variables requires modelling future homelessness as the outcome of expected trends in a set of explanatory factors.

To be more specific, simulation models estimate future levels of homelessness based on projected trends in a set of personal, economic and policy variables.[1] For instance, future levels of homelessness in the medium term are estimated conditional on demographic and economic changes.

Estimated trends in explanatory variables will feed in such models to arrive at conditional estimates of future homelessness trends. Ideally, outputs from other government models should be used as data inputs to ensure that the '*one version of the truth'* principle holds.

For instance, the Office for Budget Responsibility (OBR) forecasts of economic variables are potentially an important source of data inputs for the suite of models for homelessness. Estimated medium-term trends in GDP, inflation, labour market indicators (labour market participation, employment and unemployment), income (earnings, household disposable income) and housing market indicators (housing prices, housing stock and property transactions) are included in OBR's economy forecast among other variables.

Below is a suggestive list of forecast sources and models generating outputs that could feed in the homelessness simulation models:

- Reading-CLG Affordability model (for house prices, household formation, affordability outcomes under different supply scenarios),[2]
- IFS tax and benefits micro-simulation model – TAXBEN,[3]
- IFS projections of median income, inequality and poverty,[4] and
- ONS population projections (e.g. total population, gender and age structures, migration, population across regions).[5]

When forecasts of homelessness predictive factors are not available, administrative data can be used to predict future values of the variables of interest. Simple time-series equations can be included in the homelessness model suite to estimate trend-based projections of homelessness predictors. These equations will be fitted to time-series data issued by government sources. For example, MHCLG statistics on house building and new planning applications[6,7] as well as Continuous Recording (CORE) statistics for social housing lettings[8] can be used to arrive at estimates of future trends in a set of variables that are associated with homelessness.

Notes

[1] See the "Review of models of homelessness" for a more detailed discussion of how the simulation models work.

[2] See here for an overview of the Affordability model: https://webarchive.nationalarchives.gov.uk/20121029114150/http:/www.communities.gov.uk/documents/507390/pdf/1345079.pdf

[3] See here for a description of TAXBEN: https://www.ifs.org.uk/publications/572

[4] IFS model description and key findings on poverty and inequality can be found here: https://www.ifs.org.uk/publications/10028

[5] See here for ONS population projections: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections

[6] https://www.gov.uk/government/statistical-data-sets/live-tables-on-house-building

[7] https://www.gov.uk/government/statistical-data-sets/live-tables-on-planning-application-statistics

[8] https://www.gov.uk/government/collections/rents-lettings-and-tenancies

A.2 Upcoming homelessness data collections

A.2.1 Rough Sleeping Initiative (RSI)

The Rough Sleeping Initiative (RSI), originally introduced in 1990, is currently part of the wider objective to halve rough sleeping by 2022 and eliminate it by 2027.[43] It was initially developed outside of legislation to accommodate the needs of homeless people who fell out of the statutory homelessness definition. 83 LAs in England have recently secured MHCLG funding to support and provide accommodation to people who live on the streets.

The LAs that participate in RSI report a set of detailed data on rough sleeping on a monthly basis. The LAs also conduct street counts of rough sleeping groups three times a year (September, January, and March). The counts are further disaggregated on the basis of the same demographic characteristics reported in the official statistics. However, the counts are not part of the official statistics comprising annual counts or estimates of rough sleeping groups across LA areas and are not verified by Homelessness Link.

Specifically, the following pieces of management information are reported:

- *counts of rough sleeping populations* (flows, stock levels, returns to rough sleeping and people without local connection),
- *rough sleeping out-flows* (total number of people who sleep rough who were relieved following the intervention – in emergency, temporary, long-term accommodation or through other non-housing interventions),
- *counts of people who were previously at risk and were prevented from facing rough sleeping following the intervention*, and
- *number of spaces that are funded by RSI programmes to accommodate rough sleeping groups* (emergency, temporary or long-term accommodation).

A.2.2 Rough Sleeping Evaluation Questionnaire (RSEQ) – Complex needs

MHCLG recently introduced a strategy for collection of data from people who sleep rough in England. Individual-level data is collected from users of rough sleeping services funded by the Rough Sleeping Social Impact Bond (SIB) or the Rough Sleeping Grant (RSG) programme.

MHCLG has committed a £20 million Rough Sleeping Grant, which has been used to fund 48 projects across London and 97 LAs in the rest of

[43] see here for more information on the new government initiative to reduce rough sleeping https://www.gov.uk/government/news/new-government-initiative-to-reduce-rough-sleeping

England. It is targeted towards new rough sleepers or those at imminent risk of sleeping rough. In addition, the Authority has committed a £11 million Social Impact Bond (SIB) targeted at supporting the most entrenched rough sleepers with complex needs and is being used to fund eight LAs across rural and urban areas.

A questionnaire (the RSEQ) is currently administered to people who use the services funded through these two programmes across 11 areas in England. Data collected using this questionnaire will be used in the Complex Needs Evaluation. The aim of this questionnaire is to improve knowledge about the pathways and the needs of individuals with complex needs who sleep rough, while the research aims to explore what works best in supporting them and preventing rough sleeping, taking into consideration local context and funding mechanism.

Service users will first be asked to answer a baseline questionnaire which will be followed by two shorter follow-up questionnaires over the course of a year. These aim to track participants' outcomes following receipt of rough sleeping services.

The baseline questionnaire collects information on the following broad areas: participants' demographic characteristics, housing situation (homelessness/settled home), short-term and temporary accommodation, sofa surfing, rough sleeping, supported housing, asking LAs for help and advice, health and social support, other support needs, and schooling and institutional history. In particular, the following variables are included:

- *demographic characteristics of people who are sleeping rough* (gender, ethnicity, nationality, sexual orientation),
- *homelessness experiences* (types of accommodation (long-term/short-term) last night and during past month and length of stay. Housing owned by the service user or family and partners, social rented housing, supported housing are considered long-term accommodation. Temporary accommodation, hostels and refuges, emergency accommodation such as shelters and B&B hotels, move-on accommodation, tents, squats and sofa surfing are considered short-term accommodation),
- *feeling at risk of homelessness,*[44]
- *housing-related issues during the past month* (warning letter, threat of eviction, nuisance/anti-social behaviour complaint, suspension or sanctions from benefits),
- *short-term reasons for leaving last secure, long-term accommodation* (end of housing contract or notice from landlord, relationship breakdown, domestic violence)
- *past or current experiences of temporary accommodation, rough*

[44] The following question is included in the questionnaire: 'Do you feel secure where you are living or do you feel at risk of homelessness?'

sleeping, sofa surfing and supported housing (timing and length of the events, number of times they have occurred for sofa surfing and rough sleeping, age they first experienced event)

- *contact with housing and other services for support* (number of times users contacted local authorities because they were homeless or at risk of homelessness, types of assistance received, e.g. accommodation, advice, etc., non-housing related organisations they were in touch with),
- *health and other support needs* (self-rated health, long-standing illness or disability, mental and physical health support or drug and alcohol treatments, drug and alcohol misuse),
- *thoughts and feelings* (relaxed, optimistic, useful, close to other people, dealing with problems well, thinking clearly, close to family), and *social support*
- *institutional history* (exclusion from school, time in care, prison, armed forces),
- *income from work* (if worked, current or past), and
- *welfare benefits* (amount).

The follow-up questionnaires ask service users about their recent homelessness and housing conditions, use of homelessness and other welfare services and substance abuse.

Data collected using the questionnaire will potentially be linked to administrative data on public services use and benefits take-up, including health care services, substance use treatment programmes, and housing and other benefits in order to assess overall costs of rough sleeping.

A.2.3 Rough Sleeping Evaluation Questionnaire (RSEQ) – Costs of Homelessness

A slightly different version of the RSEQ will serve as the main data collection tool for measuring overall costs associated with homelessness and rough sleeping. The data will be analysed in the context of the 'Cost of Homelessness' project that aims to measure overall costs associated with rough sleeping.

LAs across 10-15[45] areas in England that offer rough sleeping services funded through the Rough Sleeping Initiative (RSI) and Rough Sleeping Grant (RSG) programmes will be part of the 'Cost of Homelessness' research.[46] Users of the services that will participate in the study will be asked to fill in the version of RSEQ by the service providers. The project will

[45] The final number of areas has yet to be decided.

[46] The areas that will participate in the 'Cost of Homelessness' research might overlap with but not be exactly the same as areas surveyed in the context of the 'Complex Needs' project.

be completed in two or three waves, each collecting cross-sectional data from services users. There is scope for data linkage to facilitate further analysis of individual experiences of rough sleeping and service effectiveness.

Data on service use will be measured against the Unit Cost database to estimate overall costs of services and benefits associated with homelessness and rough sleeping. Moreover, linking data to other datasets held by government departments and agencies – health care use, drug and alcohol treatment, statutory homelessness applications, welfare benefits and the criminal justice system – will allow for a more detailed assessment of costs related to homelessness and rough sleeping.

Ministry of Housing,
Communities &
Local Government

Department
for Work &
Pensions

Homelessness

Causes of Homelessness and Rough Sleeping

Feasibility study

alma economics