



# Application of predictive analytics in social care: tools to support decision making

Professor Vania Sena

Anna Crispe, Seb Smith

Dr Olga Sergushova, Liviu Andrei Toderas

Catalyst Conference 26 June 2019 #UoECatalystConf

Catalyst is funded by:



University of Essex Catalyst project working in partnership with:





# Catalyst Project - Funding

- How can Universities enable innovative, far-reaching responses to pressing national and local issues leveraging University expertise
- Working in partnership with Essex County Council and Suffolk County Council
- £4M project running from 2016 -2019 :
  - £2.2M from Higher Education Funding Council of England **Universities as 'Anchor' Initiative**
  - Co-funding from University of Essex
  - Co-funding from both Councils' Transformation Challenge Awards (TCA)



# Catalyst Project

## 3 work streams:

- **Risk Stratification:** Development of smart analytical tools to predict risk of community members becoming vulnerable and targeting services provision where it is most needed
- **Evaluation:** Development of evaluation tools to provide evidence based impact of public services; 'Spotlight Toolkit'
- **Essex Volunteering Hub:** strengthening the University's partnership with community-based volunteering hubs

## Outcomes:

- Unique opportunity to apply data analytics and evaluation expertise to real life problems
- Provision of scalability and bespoke analysis tools supporting the Council's daily decision making
- Establishing working partnerships between academic researchers and public services colleagues for public benefit



# Suffolk County Council Perspective

- 40% budget cuts over last five years – but need / complexity rising
- Highly exploratory – open minded about whether it would work at all but felt we should explore
- Fantastic access to multi-disciplinary people, skills and resources
- Focus on supporting practitioners to assess risk
- Focus on early intervention
- **Not** currently using the tools you will hear about – we need to do more work about the ethics of use, and on how to safely integrate algorithmic risk assessment into practitioners' decision making, in a way that is acceptable to the families we work with



# Overview

- What is Predictive Modelling? A short introduction
- Collaboration with Suffolk County Council (SCC)
  - Children at risk of entering care (CiC)
  - Multi-Agency Safeguarding Hub (MASH)
  - Platforms
  - Further Steps



# What is Predictive Modelling?

- Predictive modelling is a set of tools used to analyse current and historical data to make predictions about future or otherwise unknown events
- Predictive models use one or more classifiers (variables) in trying to determine the probability of an outcome
- Models usually learn from a training set of historical data and then make predictions on a testing set of current data
- Ideally, modelling is done in order to identify opportunities for intervention before the occurrence of the adverse outcome



## **Absolute risk example:**

- Among the 1,213 youth in the study population who aged out of foster care, 335 experienced homelessness at some point over the next 12 months
- Thus, one in four young people experienced homelessness after aging out of care
- Usually expressed as rate (1/4), or percentage 25%, or probability 0.25

## **Including risk factors**

- Where youth live and how much they move while in placement matters
- Being a parent may increase/decrease probability of a youth experiencing homelessness

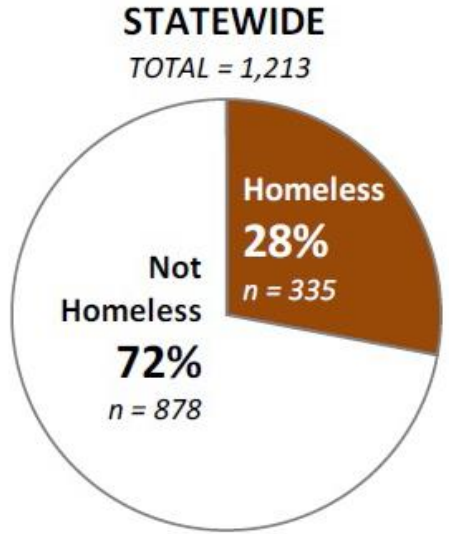
# Risk

		Youth parent		Total
		Yes	No	
Homeless	Yes	192	143	335
	No	336	542	878
Total		528	685	1213

- Probability of being homeless:  $P(H) = 335/1213 = 0.276$
- Probability of being homeless given that youths is a parent:

$P(H/Y) = P(H \text{ and } Y)/P(Y) = 192 \text{ out of } 528 \text{ out of } 1213$   
 $P(H/Y) = 0.364, P(H/ \text{not } Y) = 0.209$

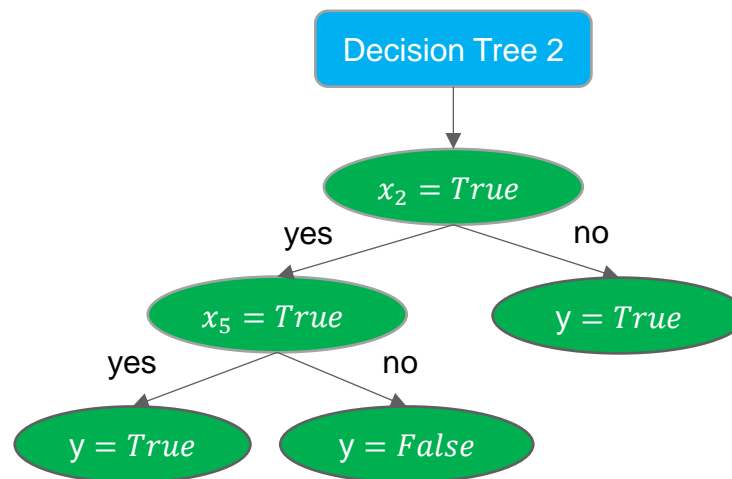
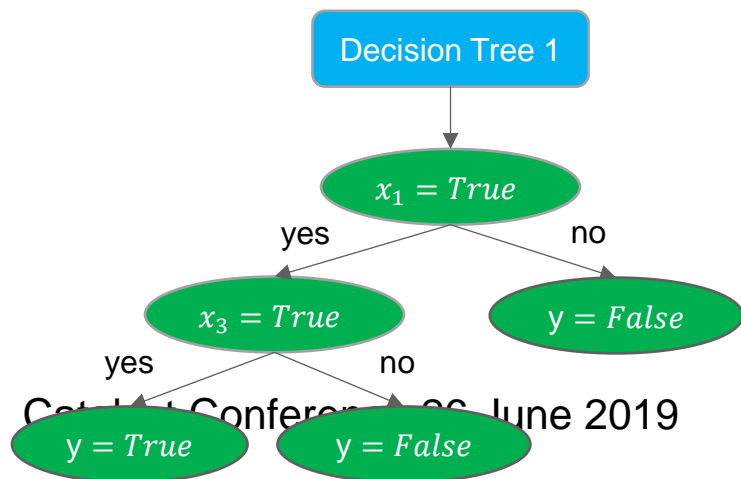
- It is more likely to be homeless if the youth was a parent
- There is a 15.5% increase at the absolute risk, if parent
- The risk of being homeless when parent is 1.74 times higher than what another youth with similar characteristics would experience if they were not parents





# Methods

- **Linear Regression** sees the outcome variable as a linear combination of the regression parameters
- **Logistic Regression** is used when the outcome variable is binary or if we want to measure the probability of the event
- **Decision tree** – a decision support tool that uses a tree-like structure where each node represents a test on an attribute, each branch is the outcome of the test, and each leaf corresponds to an outcome (decision taken after considering all attributes)
- **Random Forest** – a machine learning method that is trained on data by constructing multiple decision trees (typically 500) and then makes a prediction as an average of outcomes from each tree



Random Forest





# **Children at Risk of Entering Care**



# Who are “Children in Need”?

**Children in need** are defined in law as children who are aged under 18 and:

- Need local authority services to achieve or maintain a reasonable standard of health or development
- Need local authority services to prevent significant or further harm to health or development
- Are disabled

**What is a child protection plan?** A child protection plan is a plan drawn up by the local authority. It sets out how the child can be kept safe, how things can be made better for the family and what support they will need.



# Children in Care

A child subject to a court order child or a child accommodated under a voluntary agreement for more than 24 hours is known as a looked after child or child in care. Looked after children are typically:

- Living with foster parents
- Living in a residential children's home or
- Living in residential settings like schools or secure units

A child stops being looked after when they are adopted, return home or turn 18.

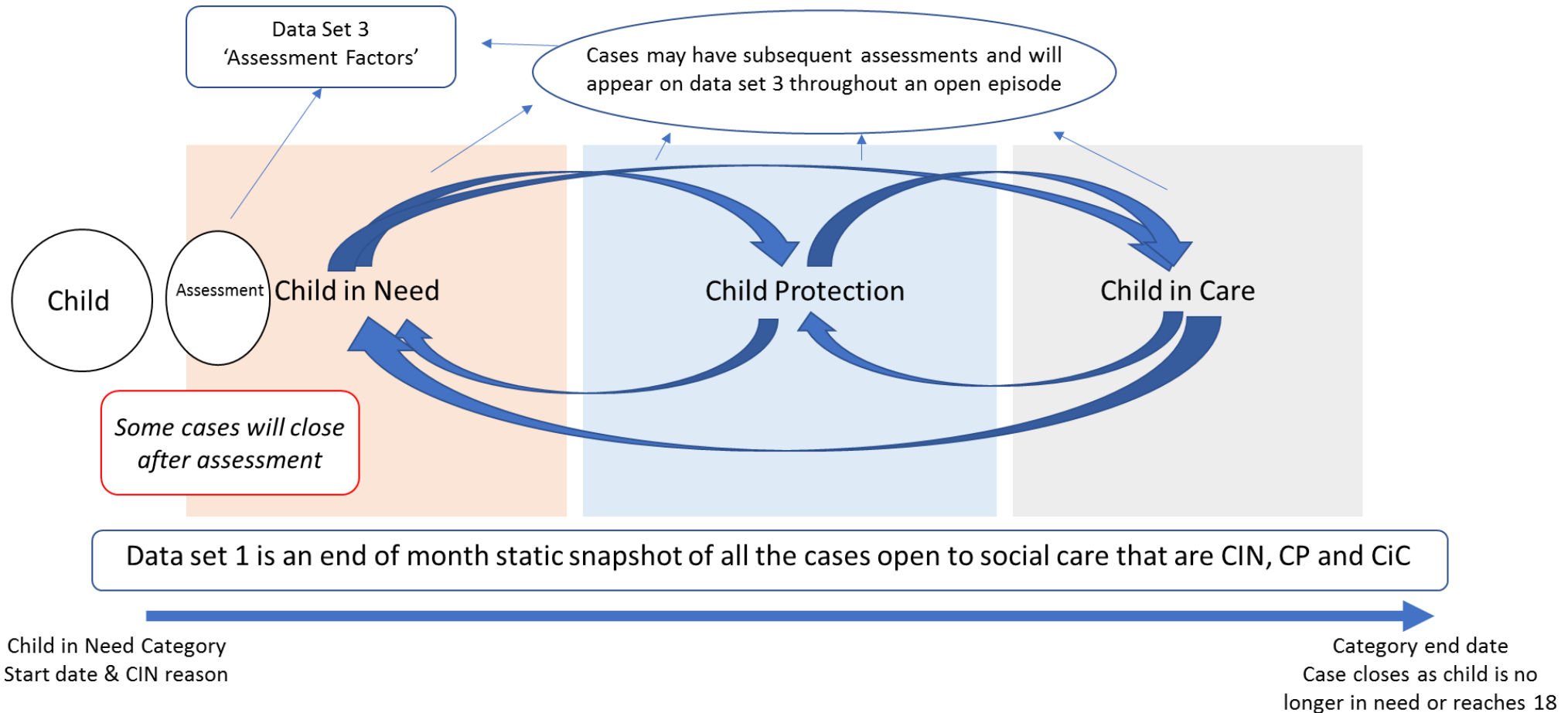
If a child return home, they may remain a child in need or on a protection plan until the local authority deems that they are no longer in need or require a plan set out for additional support.



# Case Flow

- An “episode” is the time period when a child may be open to social care services as either a Child In Need (CIN), Child subject to a Child Protection Plan (CP) or Child in Care (CiC)
- Each episode has a category start and end date (where there is no end date the episode is still open)
- A child may progress through categories of CIN to CP or CiC if the risks increase. If risks decrease, the child may progress from CiC to CP or CIN. The child could also skip stages going from CIN straight to CiC for example.
- The assessment factors are started to be collected shortly after the episode start date. Assessments provide information about identified risk factors, such as concerns about alcohol abuse, domestic violence, mental and physical health, drug misuse etc.

# Case-Flow Diagram





- Dataset 1 contains anonymised basic information about “episodes” collected from January 1995 up to March 2018
- It consists of 44472 (2626 CiC) observations of 22 variables
- Dataset 2 is considered as an addition to Set 1 and provides information about risk factors identified during assessments
- It consists of 20185 (868 CiC) observations and 53 variables
- Dataset 2 has records starting from year 2014, thus, not all observations from Dataset 1 can be matched with Set 2





# Model Performance

- The model assess probability of each episode to be in a care category
- In other words, it provides a risk level for a child to progress to care at the considered stage
- 6% of all dataset's entries corresponds to "Children in Care" episodes
- Basic Information Model
  - Accuracy – 76%      Sensitivity – 78%      Specificity – 74%
- Basic Information and Assessment Model
  - Accuracy – 83%      Sensitivity – 87%      Specificity – 80%



# Outcomes

- ✓ The model is currently in the “proof of concept” stage and going through the tests on historical data
- ✓ The output can be generated, for instance, as a ranking of cases based on predicted risk level
- ✓ It can be used to prioritise the workload of workers and identify cases that due to subtle characteristics might not have gained much prior attention
- ✓ Labelling a case as CiC does not mean that it is going to end up as CiC, it means that it has a higher likelihood than the rest of the cases
- ✓ The model should not direct decision making, but be used as an additional tool to support the process



# **Multi-Agency Safeguarding Hub (MASH)**



# MASH Overview

- MASH services can be found in multiple county councils
- Made up of several organisations who are responsible for safeguarding adults and children
- E.g. SCC, Police, Health Services, District and Borough councils, Education, Probation and Youth Offending Services
- MASH has been developed from a need to have effective information sharing between relevant agencies
- MASH supports principles of 'Working Together 2018'



## Data

A large number of entries were available on MASH systems, easily to extract, offered an opportunity to understand and use risk factors for predictions

- Data available directly from MASH systems
- 148000 entries and 270 variables from 2015 to 2018
- Data was pseudonymised (each individual has a unique ID number, traceable only by MASH team)

## Data Cleaning

After checking the reliability of the data (missing values, duplicated records, misspelled values etc.) we were left with

- ~58000 out of ~148000 entries
- ~25000 are unique individuals
- 71 variables out of 270

# Cleaning “Outcome” variable

- Balancing the outcome and re-coding

BEFORE							AFTER	
ICS	ADIFSOC	AS	DCYP AS	DOT	EHCYP	EHOTH	MASHED	NOT MASHED
7959	25	6	81	13	1189	610	28313	29894
IAG	ISGLDO	NOT MASHED	PFA	R	S47	S47s		
14675	19	29894	3	1767	1955	10		



# Model performance

- Current model can predict with 74% accuracy if a child should go through a MASH review or not
- We can present these results in various ways:
  - We can rank based on percentage
  - Show percentage next to entry
  - Highlight only important cases

Probability to MASH	
Risk Rank	Number in sample
Very Low (under 25%)	6104
Low (25-50%)	1323
Moderate(50 -75%)	2344
High(over 75%)	6144



# Outcome

- Live data is compared to the existing model and a risk level output is provided for each entry
- Example of two risk profiles:
  - ❖ Male, 15, IP14, two siblings, three adults, parents aware, family dysfunction, current adult drug use, multiple risk, previously underwent MASH review 1 – 71% probability to MASH
  - ❖ Female, 15 , IP11, two sibling, one adult, antisocial behaviour, current alcohol use, themselves as risk, previously underwent MASH review 5 – 40% probability to MASH
- As a result data recording policies are changing





# MASH Perspective

- MASH had a large number of contacts many of which do not require a MASH assessment
- Based on historic data we want to identify children more likely to go through for a MASH assessment
- Help MASH provide timely interventions to children deemed at risk
- Help with training new professionals joining the MASH team



# MASH Perspective

- Focus limited social work resource on those children most at risk
- Important to retain confidence of partners
- Not a cost cutting exercise
- Utilises the best historic decision making



# Platform

- Created in order to provide MASH and Social Care workers easy access to the outputs of the complex algorithms used
- These can also include other information, like graphs and descriptive statistics
- The platforms are a good way to engage people with little R knowledge and enable them to use the results
- We provide training in order to keep the methods used transparent and give the opportunity to professionals to scrutinize and change the algorithms



# Further steps

- Pilot MASH tool for ~2 months
- Provide training on how the tool/model works (30<sup>th</sup> of May)
- Look into ways to incorporate the platform (Nuffield application)
- Evaluate the benefits brought by the projects
- Work with the departments to further improve the platforms



# Questions



## Evaluation

Empowering public services to evaluate the impact of their work

## Risk Stratification

Using predictive analytics to anticipate those at risk and to better target resources

## Essex Volunteering Hub

Providing benefits to local community and students through volunteering

---

### Contact us:

E [catalyst@essex.ac.uk](mailto:catalyst@essex.ac.uk)

T +44 (0)1206 872057

[essex.ac.uk/research-projects/catalyst](https://essex.ac.uk/research-projects/catalyst)

 @HEFCE\_Catalyst #UoECatalystConf

