PHD RESEARCH · UNIVERSITY OF BRISTOL

Who gets lost on the way to university?

Tracing care-experienced young people from GCSEs to a degree, with machine learning

Joanna Papageorgiou

Who enters higher education by age 22

14%

Care leavers

56%

General population

With low or no GCSE grades: **4% or less** of care-experienced young people entered HE, against **28%** of the general population.

Source: Feinstein et al. (2025), TASO/REES. England birth cohort 1998/99, n = 556,240. General population excludes young people with social care contact or free school meals.

Three care leavers, one recorded outcome

STUDENT A

Good GCSEs. Never applies through UCAS.

Lost at: applying

STUDENT B

Applies to four universities with the right grades. No offers.

Lost at: offers

STUDENT C

Accepts an offer. Never enrols.

Lost at: enrolment

A single yes/no outcome records all three as *“did not enter higher education”*. *education”*.

The pipeline: nine stages from GCSEs to a degree

S0	S1	S2	S3	S4	S5	S6	S7	S8
GCSEs	**Level 3**	**Apply**	**Choices**	**Offer**	**Placed**	**Enrol**	**Year 2**	**Degree**

Green: school and college. Yellow: applying. Orange: university. The dashed box describes how people apply.

> **At each stage:** of the people who reached it, what share moved on? That is the **transition rate**.
>
> Multiplied together, the stages show where the difference in progression builds up.

New: linked data like this has only become available for this research in the last year or two, following Feinstein et al. (2025).

NPD · school records

CLA and CIN · care records

UCAS · applications at 18

HESA · enrolment and degrees

Linked through **LEO** in the ONS Secure Research Service

At every stage: two models, two checks

MODEL 1

Logistic regression

The standard method. Average effects, directly comparable with Feinstein et al.

MODEL 2

CatBoost

Machine learning with many small decision trees. Finds interactions, such as care status × grades × local authority.

CHECK 1

SHAP explanations

Shows which factors push each young person's predicted chance up or down.

CHECK 2

Fairness diagnostics

Tests whether the model is equally accurate across subgroups within the care-experienced group.

What the research will show

1

Where in the pipeline the difference in progression opens up, and how much at each stage.

2

Which factors matter at each stage, and for which young people.

3

Evidence that Virtual School Heads, local authorities and universities can act on, stage by stage.

Thank you. Questions.

Joanna Papageorgiou · University of Bristol