

Mapping and evaluating the use of contextual data in undergraduate admissions in Scotland

An Impact for Access project funded by the Scottish Funding Council

Vikki Boliver, Stephen Gorard, Mandy Powell, and Tiago Moreira
Durham University

October 2017

Report 3

Evaluating the validity and reliability of potential indicators of contextual disadvantage

Evaluating the validity and reliability of potential indicators of contextual disadvantage

Overview

This report provides a conceptual evaluation of the validity (or appropriateness) and the reliability (or trustworthiness) of a range of potential indicators of contextual disadvantage. The aim of the report is to develop some initial recommendations regarding which potential indicators are likely to be suitable for use by universities for the purposes of contextual admissions. We distinguish between those indicators that are suitable for use alone on an EITHER/OR basis; those that are not suitable for use alone but which may be suitable for use in combination on an AND/ALSO basis; and those that are unsuitable for use as indicators as contextual disadvantage even in combination, at least in their current form.

Section 1 discusses what is meant by validity and by reliability and outlines why it is crucial to select valid and reliable indicators of contextual disadvantage if contextualised admissions policies are to be fully effective and not potentially counterproductive. We distinguish between two important considerations when evaluating validity and reliability: (1) evaluation of the risk that an indicator may identify an applicant as contextually disadvantaged when they are not (known as a 'false positive'), and (2) evaluation of the risk that an indicator may identify an applicant as not contextually disadvantaged although they are (known as a 'false negative'). We highlight that, while the most valid and reliable indicators are of course those which yield few false positives *and* few false negatives, indicators which yield few false positives can be considered suitable for use alone on an EITHER/OR basis regardless of the number of false negatives they yield (and in fact using multiple indicators of this type on an EITHER/OR basis will tend to reduce false negatives). We also show that indicators which yield more than a few false positives may be suitable markers of contextual disadvantage if used in combination on an AND/ALSO basis.

Section 2 reports on our conceptual evaluation of the validity and reliability of a range of potential indicators of contextual disadvantage. We assess each of the contextual indicators currently being used by Scottish universities to inform admissions decisions (see Report 1), as well as several other indicators not currently widely used in Scotland but in use by universities across the rest of the UK. We pay particular attention to the extent to which each indicator is likely to minimise false negatives and false positives. We recommend that:

- the following indicators can be considered ***suitable for use alone on an EITHER/OR basis*** on grounds that they are individual-level markers of socioeconomic disadvantage which can be administratively verified: having spent time in care, being a carer for a family member, being a refugee or asylum seeker, having been in receipt of free school meals, or being in receipt of an Education Maintenance Allowance. All of the above indicators are suitable for use on an EITHER/OR basis (i.e. applicants can be considered contextually disadvantaged if they meet any one of these criteria).

- the following indicators are ***not suitable for use alone but may be suitable for use in combination on an AND/ALSO basis***: residing in an SIMD20/40 postcode area, residing in a POLAR quintile 1 or 2 postcode area, being resident in an ACORN 4 or 5 postcode, attending a school where a high percentage of students receive free school meals or an Education Maintenance Allowance, attending a low attainment school, and attending a low progression school. These indicators are not suitable for use singly on grounds that they may not accurately capture the personal circumstances of specific individuals; however, they may be suitable for use on an AND/ALSO basis (i.e. applicants may be considered contextually disadvantaged provided they meet, for example, any two of these criteria).

- the following indicators are ***not suitable for use alone or in combination***, at least in their current form: parental education and parental occupational social class, because they are not administratively verifiable; and living in a rural area and attending a non-selective non-fee-paying school, because these categories encompass a large number of individuals who are not personally socioeconomically disadvantaged.

1. Key considerations in the assessment of validity and reliability

1.1 The meaning and importance of validity and reliability

The effectiveness of any contextualised approach to admissions depends first and foremost on the validity and reliability of the indicators used to identify contextually disadvantaged applicants.

Assessing the validity of indicators of contextual disadvantage involves making judgements about their *appropriateness*; about the degree of precision with which they capture the thing they are intended to capture. Assessing reliability, on the other hand, involves judgements about the *trustworthiness* of instruments and data sources used for measurement; about the degree of accuracy and consistency with which indicators measure the thing they are intended to measure.

More specifically in relation to validity, an indicator that is too narrowly defined will result in some applicants being identified as not contextually disadvantaged even though they are in reality (false negatives), whereas an indicator that is too broadly defined will result in some applicants being identified as contextually disadvantaged when they are not (false positives). Validity concerns are relevant regardless of whether the indicator is measured at the individual, area or school level, but tend to be particularly pronounced for area-level and school-level indicators since these may be poor proxies for the characteristics of individuals (a problem known as the ecological fallacy), and even of the wider environment in which an individual lives or was educated if based on a large number or diverse set of households or pupils.

With respect to reliability, an indicator will produce a significant number of false negatives and/or false positives if the measurement instrument produces inconsistent results across cases. Threats to reliability are most pronounced where the information used is self-reported and unverified since the reporter may give incorrect information unintentionally or otherwise. The reliability of an indicator is also reduced in proportion to the amount of missing data. These threats to reliability are relevant to all potential indicators of contextual disadvantage regardless of whether they are measured at the level of the individual, area or school.

The validity of an indicator and its reliability are both crucial. If an indicator is valid but not reliable, or if it is reliable but not valid, it is not safe to use. It is important to appreciate, however, that it is highly unlikely that any single indicator will be entirely valid and entirely reliable. As such, the task is to identify indicators that are sufficiently valid and sufficiently reliable, either on their own or when deployed in combination.

1.2 Distinguishing between false negatives and false positives

Assessing the validity and reliability of contextual indicators involves assessing the extent to which they yield false negatives and false positives, alone and in combination. Minimising the incidence of false negatives and minimising the incidence of false positives are both

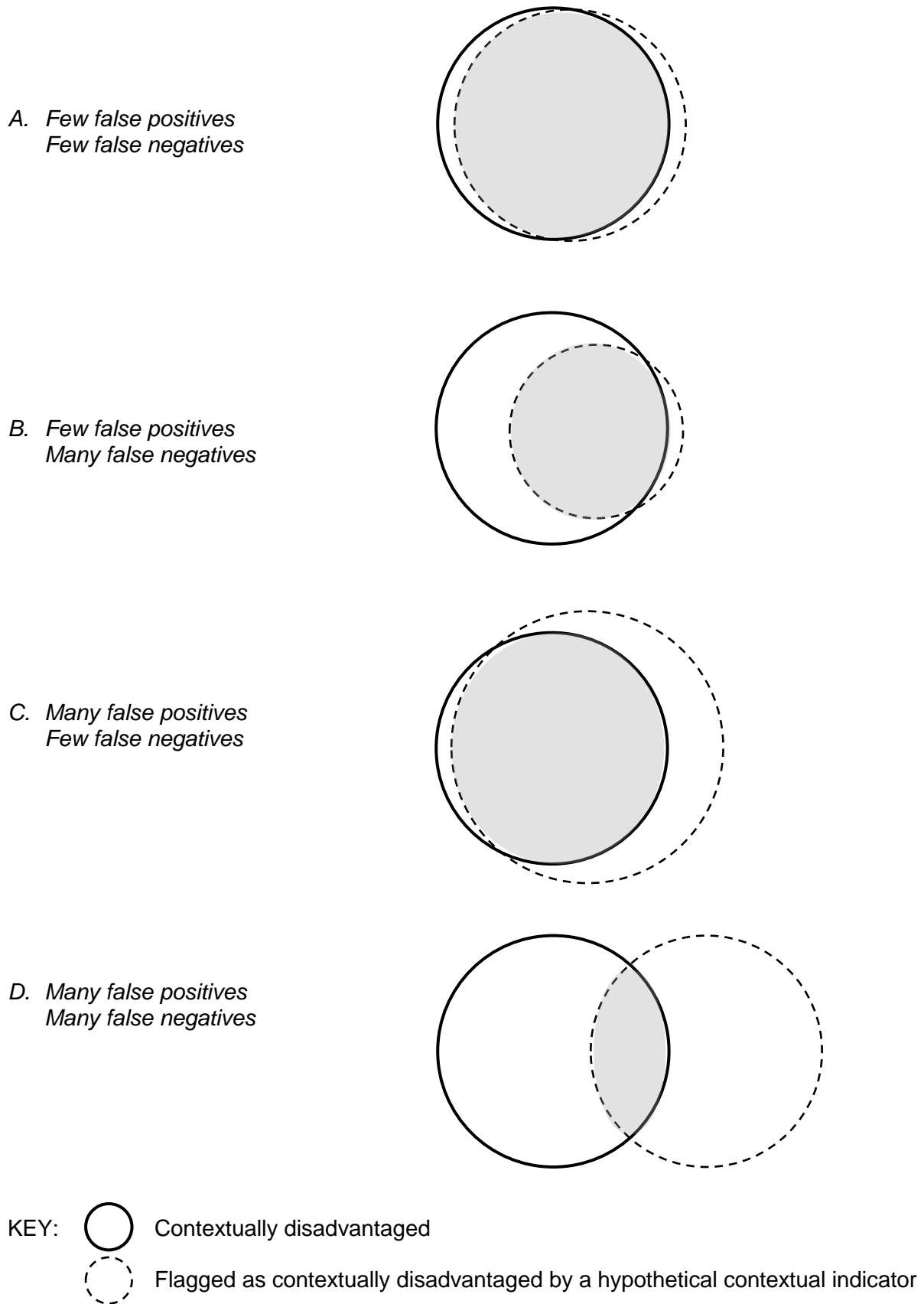
important objectives. However, minimising the incidence of false positives is more important than minimising the incidence of false negatives. This is because an indicator which produces a significant number of false negatives may compromise the effectiveness of contextualised admissions given that a significant number of intended beneficiaries will not be reached; but while this means that the indicator does not do as much good as it might, it at least does no harm relative to the *status quo*. In contrast, an indicator which produces a significant number of false positives, thereby benefitting a significant number whose actual circumstances do not warrant it, may in fact be counterproductive; if the number of false positives outweighs the number of true positives, the overall result will be more harm done than good relative to the *status quo*. This distinction between false positives and false negatives, and the greater importance of minimising the former than of minimising the latter, is illustrated Figure 1, below.

In each of the four diagrams labelled A, B, C and D in figure 1, the thick-lined circle represents those who are in fact disadvantaged, while the dashed-lined circle represents those flagged as disadvantaged by a hypothetical contextual indicator. The intersection of these two circles, shaded in grey, represents those who are in fact disadvantaged and have been *correctly* flagged as such by the contextual indicator (i.e. true positives). The unshaded portion of the thick-lined circle represents those who have been incorrectly flagged as not disadvantaged even though they in fact are (false negatives), while the unshaded portion of the dashed-lined circle represents those who have been incorrectly flagged as disadvantaged though they are in fact not (false positives).

Diagram A illustrates something close to the best-case scenario in which an indicator is near-perfect in terms of its validity and reliability, producing few false negatives and few false positives. If such an indicator existed it would be highly suitable for use alone, and no other indicators would be needed.

Diagram D, at the other pole, illustrates something close to the worst-case scenario in which an indicator is highly imperfect, producing many false negatives and many false positives; an indicator like this clearly should not be used.

Figure 1. The distinction between false positives and false negatives



Diagrams B and C illustrate two in-between scenarios. Diagram B illustrates an indicator that is highly imperfect in producing many false negatives but is near-perfect in producing few false positives. Conversely, diagram C illustrates an indicator that is near-perfect in producing few false negatives but is highly imperfect in producing many false positives. The indicators in diagrams B and C both constitute imperfect indicators, but it is important to appreciate that the implications are not the same for both.

Because the indicator in diagram B produces few false positives, it would not be counterproductive to use it on its own (i.e. its use would not inadvertently benefit many whose circumstances do not warrant it), but because it produces many false negatives its use would compromise the effectiveness of contextualised admissions *if it was the only indicator used*. As such, indicators like that illustrated in diagram B are suitable to use singly but for contextualised admissions to be maximally effective *all such* indicators should be made use of on an EITHER/OR basis. In contrast, although the indicator in diagram C produces few false negatives (i.e. it captures most of the intended beneficiaries), the fact that it produces many false positives means that it would likely be counterproductive to use it on its own. However, indicators like those in diagram C may be suitable for use in combination on an AND/ALSO basis.

1.3 Distinguishing between suitability for use on an EITHER/OR vs. an AND/ALSO basis

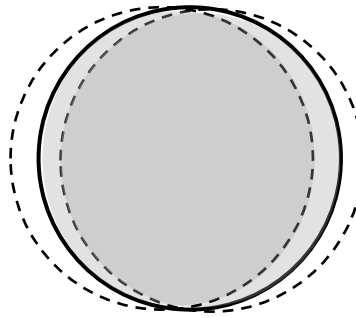
Figure 2 develops the points made above. Diagram A in figure 2 demonstrates how two indicators of the kind identified in diagram A of figure 1 may be suitable for use on an EITHER/OR basis. Each of the indicators in diagram A of figure 2 produce few false positives and few false negatives when used alone, and when both are used on an EITHER/OR basis the overall incidence of false negatives is likely to be reduced further. It is of course unlikely that there is one, let alone two, such indicators which produce few false positives and few false negatives; at best we are likely to find indicators which produce few false positives but many false negatives (Diagram B in figures 1 and 2).

Diagram B in figure 2 demonstrates how two indicators of the kind identified in diagram B of figure 1 are likely to be suitable for use on an EITHER/OR basis. Each of the indicators in diagram B of figure 2 produce few false positives, but many false negatives when used alone. However, when both are used on an EITHER/OR basis, the incidence of false negatives is reduced overall.

Figure 2. The distinction between use on an EITHER/OR vs. and AND/ALSO basis

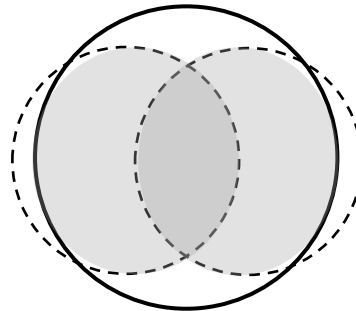
A. *Few false positives*
Few false negatives

Suitable for use alone
on an EITHER/OR basis



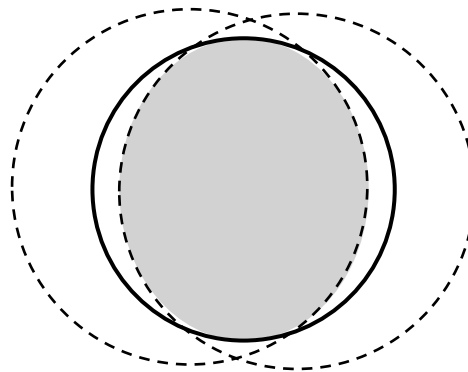
B. *Few false positives*
Many false negatives

Suitable for use alone
on an EITHER/OR basis



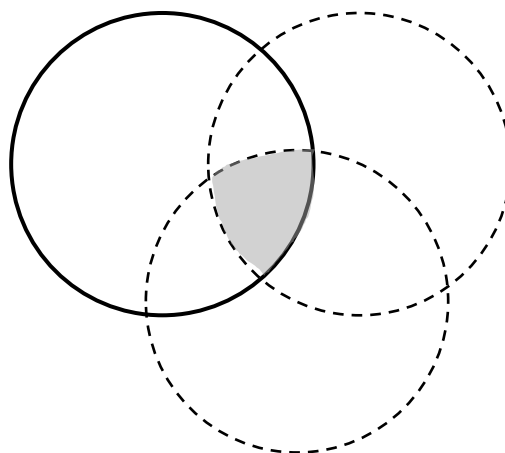
C. *Many false positives*
Few false negatives

May be suitable
for use in combination
on an AND/ALSO basis



D. *Many false positives*
Many false negatives

Unsuitable for use
alone or in combination





KEY:  Contextually disadvantaged
 Flagged as contextually disadvantaged by a hypothetical contextual indicator

Diagram C in figure 2 demonstrates how two indicators of the kind identified in diagram C of figure 1, though not suitable for use alone on an EITHER/OR basis, may be suitable for use in combination on an AND/ALSO basis. Each of the indicators in diagram C of figure 2 produce few false negatives but many false positives; however, their use in combination on an AND/ALSO basis could result in few false positives overall if the two indicators are only moderately correlated with one another (i.e. if they overlap, but only partially) as is the case in diagram C of figure 2. Conversely, if the two indicators are strongly correlated with one another, the incidence of false positives will be larger.

Diagram D in figure 2 shows how two indicators of the kind identified in diagram D in figure 1, are unlikely to be suitable for use even in combination on an AND/ALSO basis. Each of the indicators in diagram D of figure 2 produce many false negatives and many false positives; if the two indicators were used in combination on an AND/ALSO basis, the number of false negatives would increase, and the number of false positives, though reduced, would remain substantial.

Section 2 of this report discusses which indicators are likely to be suitable for use alone on an EITHER/OR basis, which may be suitable for use in combination on an AND/ALSO basis, and which are unsuitable for use.

2. A conceptual evaluation of the validity and reliability of potential indicators of contextual disadvantage

This section turns to a conceptual evaluation of the validity and reliability of a range of potential indicators of contextual disadvantage, considering the degree to which each can be expected to yield false positives and/or false negatives. We focus on indicators relevant to contextual disadvantage defined as the experience of or exposure to challenging socioeconomic circumstances considered likely to have impacted negatively on educational attainment and/or progression. This definition of disadvantage as challenging socioeconomic circumstances includes a range of economic, social or cultural forms of absolute or relative deprivation which stem from inequalities in the wider society. The definition of context as experience of or exposure to these forms of disadvantage encompasses the characteristics and circumstances of individuals and/or their immediate households, as well as the local environment in which individuals live or go to school.

Our focus on socioeconomic forms of disadvantage means that we do not consider here any demographic classifications which have been identified as ‘protected characteristics’ under Equalities Law, nor do we consider ‘extenuating personal circumstances’, even though these are mentioned by some institutions in their contextual admissions policy documents (see Reports 1 and 2). The protected characteristics referred to in contextualised admissions policy documents include age (referring specifically to mature applicants), disability, and broad references to ‘other protected characteristics’. The extenuating personal circumstances referred to include serious illness or close family bereavement, particularly where experienced during school examination periods. We recognise that these characteristics and circumstances may impact negatively on educational attainment and progression and that as such it is appropriate for universities to be cognisant of them. However, we do not consider them in this report principally because they do not constitute indicators of *socioeconomic* disadvantage. In relation to protected characteristics specifically, we would note that it is not permissible under equalities law to treat applicants differently on these grounds, although encouraging applications from members of under-represented groups is a legally permitted form of positive action. In relation to extenuating personal circumstances, we would note that where it is believed that these factors have impacted negatively on educational achievement, it is possible for individuals or their schools to request that exam boards consider mitigating circumstances when marking exam scripts and other affected assessments.

We include in this section all indicators meeting the definition of contextual disadvantage described above that are currently being used by Scottish universities for contextual admissions purposes, as well as several other indicators that are being used by universities located in the rest of the UK. We consider three categories of indicator in turn: individual-level indicators, area-level indicators, and school-level indicators. Individual-level indicators refer to the circumstances of individuals and their households. Area-level and school-level indicators, on the other hand, refer to the average circumstances of individuals and households living in the same locale or attending the same school as the individual concerned; these can be considered to represent either proxies for the circumstances of individuals, or of the environment in which an individual lives or has been educated.

INDIVIDUAL-LEVEL INDICATORS

2.1 First-generation of family members to gain a degree

This indicator is available to universities via UCAS for applicants who have completed the relevant field in their online application form, and may also be available for applicants who took part in an outreach programme where this is part of the eligibility criteria for the programme.

Validity: This indicator rests on the assumption that parents who are not educated to degree level are more likely to be socioeconomically disadvantaged and less well-placed to advise and support their children's school attainment progression to higher education. Parents' educational level, and mother's educational level in particular, is known to be positively associated with school attainment.¹ This indicator might produce few false negatives since degree-educated parents are likely to be socioeconomically advantaged and are, by definition, more familiar with what it takes to successfully navigate the university admissions systems. However, this indicator is likely to produce many false positives because many prospective university students will be the first generation in their family to go to university, making this a very socioeconomically diverse category.

Reliability: Because the information used to derive this indicator is based on applicants' self-reports, and is not readily verifiable, it risks producing many false positives and false negatives. The information applicants provide may be inaccurate if young people are not aware of their parents' highest educational qualifications. Moreover, there tends to be a substantial amount of missing data.

2.2 Parents not employed in professional or equivalent occupations

The data underpinning this indicator is collected via UCAS application forms but it not currently made available to universities until after the admissions cycle has been concluded. It may also be available for applicants who took part in an outreach programme where this is part of the eligibility criteria for the programme.

¹ Dearden, L., Sylva, K. and Sibieta, L. (2011) The socio-economic gradient in early child outcomes: evidence from the Millennium Cohort Study. *Longitudinal and Life Course Studies*, 2(1):19-40.

Validity: This indicator rests on the assumption that parents who are not employed in professional-level or equivalent occupations are more likely to be socioeconomically disadvantaged and less well-placed to advise and support their children's school attainment progression to higher education. Parents' occupational social class is associated with attainment at school,² which is largely responsible, in turn, for lower rates of participation in higher education for those from lower social class backgrounds.³ As with the previous indicator, this indicator may produce few false negatives since parents in professional-level and equivalent occupations are likely to be socioeconomically advantaged and are more likely to hold a degree themselves. However, this indicator might produce some false positives since not all prospective university students whose parents are in non-professional occupations will be socioeconomically disadvantaged (including some retired, independently wealthy, trades and so on).

Reliability: Because the information used to derive this indicator is based on applicants' self-reports, and is not readily verifiable, it is likely to produce many false positives and false negatives. The information applicants provide may be inaccurate because young people often do not know exactly what jobs their parents hold. The indicator is intrinsically hard to code. This indicator tends to have a substantial amount of missing data (around 20% of students per year) which is cited by UCAS as the primary reason this information is no longer supplied to universities until the end of each admissions cycle.

2.3 Has spent time in registered care

This indicator is available to universities via UCAS for applicants who have completed this field in their online application form, and may also be available for applicants who took part in an outreach programme where this forms part of the eligibility criteria for the programme.

Validity: This indicator focuses on those who have experienced significant personal adversity and educational disruption during childhood which has resulted in them becoming a 'looked after' child within the care system. Applicants who have spent time in care may also face

² Gorard, S., Adnett, N., May, H. Slack, K., Smith, E. and Thomas, L. (2007) *Overcoming barriers to HE*. Stoke-on-Trent: Trentham Books.

³ Marcenaro-Gutierrez, O., Galindo-Rueda, F. and Vignoles, A. (2007) Who actually goes to university? *Empirical Economics*, 32(2): 335-357.

Noble, J. and Davies, P. (2009) Cultural capital as an explanation of variation in participation in higher education. *British Journal of Sociology of Education*, 30, 5, pp. 591-605.

Chowdry, H., Crawford, C., Dearden, L. Goodman, A. and Vignoles, A. (2013) Widening participation in higher education: analysis using linked administrative data. *Statistics in Society*, 176(2): 431-457.

continuing difficulties as a result of living away from birth parents with foster parents, in a children's home, or independently. Those who have been taken into care are more likely than others to come from socioeconomically disadvantaged birth families. School attainment levels and higher education progression rates are markedly poorer for young people who have spent time in registered care relative to other young people.⁴ Having spent time in care is widely considered to be a marker of disadvantage and is likely to produce few, if any, false positives. Of course, since the vast majority of young people have never been in care, this indicator will inevitably yield a large number of false negatives.

Reliability: Although this indicator is based on applicants' self-reports, it is readily verifiable with reference to administrative data and so could produce few, if any, false positives. It is not clear whether this indicator as supplied by UCAS is officially verified; ideally UCAS would undertake to ensure its veracity, but the small numbers involved mean that institutions could seek to verify claims to looked after status on a case-by-case basis. Its reliability, however, depends on applicants' willingness to disclose their looked after status on their UCAS form or elsewhere, and for this reason will inevitably produce some false negatives.

2.4 Is or has been a long-term carer for a family member

This indicator is not commonly available to universities via UCAS, but could be included as a field on UCAS forms, and may be available for applicants who took part in an outreach programme where this forms part of the eligibility criteria for the programme.

Validity: This indicator focuses on those who are likely to have experienced significant personal adversity and educational disruption as a result of a long-term responsibility during childhood and/or adolescence for meeting the care needs of an ill, disabled or otherwise troubled family member. More carers are from socioeconomically disadvantaged backgrounds and carers typically achieve less well at school and are less likely to progress to higher education than their peers (insofar as evidence is available on this). Carers almost certainly count as disadvantaged and so this indicator is likely to yield few false positives. However, as with the previous indicator, there will inevitably be many false negatives given that the vast majority of prospective students are not carers

⁴ Croll, N., Browitt, A., Anderson, M. and Hedge-Holmes, K. (2016) The University of Glasgow and West of Scotland Local Authority partners: how to engage with MD40 pupils in higher progression schools - SFC Impact for Access Project Report. Glasgow: University of Glasgow.

Reliability: Although this indicator is necessarily based on self-reports, it is potentially verifiable by schools or by carer support organisations and so might produce few, if any, false positives. However, its reliability depends on applicants' willingness to disclose their status as a carer, and so will inevitably produce some false negatives.

2.5 Is a refugee or asylum seeker

As with the previous indicator, this indicator is not commonly available to universities via UCAS but could be included as a field in the UCAS application form. It may be available for applicants who took part in an outreach programme where it forms part of the eligibility criteria for the programme.

Validity: This indicator focuses on those who are likely to have experienced significant personal adversity and educational disruption due to displacement from their previous country of residence, to which it is unsafe for them to return. Refugees and asylum seekers are not necessarily from socioeconomically disadvantaged backgrounds in their country of origin but will often be so in their country of destination due to restrictions placed on adult family members' eligibility to work and in some cases due to young people's separation from adult family members. In any case refugee or asylum seeker status is commonly associated with significant educational disruption both before and after migration. As such, this indicator can be expected to yield few false positives. However, false negatives will be widespread.

Reliability: Although this indicator is necessarily based on self-reports, it would be verifiable with reference to administrative data and so might produce few, if any, false positives. However, its reliability depends on applicants' willingness to disclose their status as a refugee or asylum seeker, and so some false negatives are likely.

2.6 Is in receipt of free school meals (FSM)

This indicator is not currently used by Scottish universities as a contextual indicator for applicants in general. However, it is frequently used as an indicator for outreach programmes and summer schools offered by universities in Scotland and the rest of the UK, and could be used for the purposes of contextual admissions more generally. The data for this indicator could be provided by applicants' schools, or by the Scottish Education Analytical Services division for applicants educated in Scotland and the Department for Education National Pupil Database for applicants educated in England, subject to appropriate data protection and data sharing arrangements being put in place.

Validity: This indicator focuses on those whose low household income qualifies them for one or more of a range of social security benefits such as Income Support, Jobseeker's Allowance, Working Tax Credit, Child Tax Credit or Universal Credit, which subsequently qualifies children in the family for free school meals. School attainment levels and higher education progression rates are markedly lower for children known to be eligible for free school meals in comparison with other children. Those eligible for free school meals are, by definition, highly socioeconomically disadvantaged, meaning that this indicator yields few false positives. However, it will yield some false negatives, since families who are only a little above the income threshold for eligibility for free school meals are likely to be similarly socioeconomically disadvantaged and may in fact have lower incomes than families who are FSM eligible after all social security transfers have been taken into account.⁵

Reliability: This indicator would be based on administratively verified data supplied by schools or by the statistics teams which support the government departments for education. As such this indicator is likely to yield few, if any, false positives. However, since the data refers to receipt of free school meals as opposed to eligibility for receipt, and since not all eligible families provide the information needed to assess their eligibility, some false negatives will occur. The rate at which those eligible for free school meals register to receive this entitlement tends to decline as school children get older.⁶ As such, the use of data in relation to *ever* in receipt of free school meals (rather than *currently* in receipt) would help to reduce false negatives.⁷ This would require data sharing agreements between institutions and the Scottish Government as data holder.

2.7 Is in receipt of an education maintenance allowance (EMA)

As with receipt of free school meals, data on receipt of Educational Maintenance Allowance is not generally used by Scottish universities for contextualised admissions in general, but is sometimes used as an indicator for widening participation programmes and summer schools, and could be used for the purposes of contextual admissions more generally. The data for this indicator could be provided by applicants' schools, or by the Scottish Education Analytical

⁵ Hobbes, G. and Vignoles, A. (2010) Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal*, 36(4): 673-690.

⁶ CoWA (2016) Commission on Widening Access - Technical paper on measures and targets - March 2016. Edinburgh: Scottish Government.

⁷ Boliver, V., Gorard, S. & Siddiqui, N. (2015). Will the Use of Contextual Indicators Make UK Higher Education Admissions Fairer? *Education Sciences*, 5(4): 306-322.

Services division for applicants educated in Scotland and its equivalents for applicants schooled in England, Wales and Northern Ireland, subject to appropriate data protection and data sharing arrangements being put in place.

Validity: As with the previous indicator, EMA focuses on those with low household incomes – currently less than around £25,000 per annum before tax – which qualifies members of the household aged 16 to 19 for an educational maintenance allowance of £30 per week to help meet the costs of continuing in education. EMA is currently only available to pupils studying in Scotland, Wales and Northern Ireland (in England there is a 16-19 Bursary with more stringent eligibility criteria). Educational attainment and progression is significantly lower for individuals from low income households than for others,⁸ and those from low income households are, by definition, highly socioeconomically disadvantaged, meaning that this indicator yields few false positives (but more than FSM which has a lower threshold). However, as with the previous indicator, this indicator will yield some false negatives, since those who are only a little above the income threshold for eligibility for EMA are likely to be similarly socioeconomically disadvantaged and perhaps more so after social security transfers are taken into account.⁹

Reliability: This indicator could be based on administratively verified data supplied by schools or by the statistics teams which support the government departments for education. As such, this indicator would produce few, if any, false positives. However, because the data typically refers to those in receipt of rather than those eligible for EMA, and does not relate to all home countries a significant number of false negatives will occur.

AREA-LEVEL INDICATORS

2.8 Resident in an SIMD20 or SIMD40 postcode area

This indicator is available to universities via UCAS for applicants living in Scotland. A similar Index of Multiple deprivation indicator is also available from the UK government data service for applicants living in England and the rest of the UK

Validity: This indicator refers to residents in areas classified as being among the 20% or 40% most deprived areas according to the Scottish Index of Multiple Deprivation (SIMD). SIMD is

⁸ Anders, J. (2012) The link between household income, university applications, and university attendance. *Fiscal Studies*, 33(2): 185-210.

⁹ Hobbes, G. and Vignoles, A. (2010) Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal*, 36(4): 673-690.

a composite indicator of relative deprivation based on 38 measures grouped into seven, differently weighted, domains (income, employment, education, health, access to services, crime, and housing). Those living in areas ranked in the lowest two quintiles on this measure (SIMD40), and particularly those living in areas ranked in the bottom quintile (SIMD20), are more likely than others to experience conditions which impact negatively on opportunities and outcomes, including educational attainment¹⁰ and progression to higher education with rates of university entry currently thirty percentage points higher for those from areas in the top as compared to the bottom SIMD quintile.¹¹ Evidence for the West of Scotland indicates that SIMD is correlated with the percentage of students in receipt of free school meals; around 34% of students in the lowest SIMD decile were receiving FSM compared to around 2.5% of students in the highest SIMD decile.¹² However, since SIMD is an area level measure which aggregates information for a large number of people living in the same locale (approximately 760 per area) it may be a poor proxy for the personal circumstances of individuals, except perhaps in specific locales.¹³ If used to identify disadvantaged individuals, SIMD20/40 can be expected to yield a non-negligible proportion of false positives (individuals living in deprived areas who are not themselves deprived) and a non-negligible proportion of false negatives (individuals who are deprived but are not living in deprived areas). For example, the supporting documentation for SIMD states that only around one in three people living in a deprived area are income deprived, and that two out of three people who are income deprived do not live in deprived areas.¹⁴ Similarly, the Commission on Widening Access Technical Paper found that just over one half of all pupils receiving free school meals do not live in SIMD20 neighbourhoods and notes that while SIMD “is currently the most suitable measure of disadvantage for the purposes of measuring progress and setting targets...it is important, however, that *decisions about individuals and the support they require* are not made using SIMD alone” (COWA 2016: 4, emphasis added).¹⁵

¹⁰ Scottish Government (2016) Introducing the Scottish Government Index of Multiple Deprivation 2016. Edinburgh: Scottish Government.

¹¹ UCAS (2015) End of Cycle Report 2015. December 2015. Cheltenham: UCAS.

¹² Croll, N., Browitt, A., Anderson, M. and Hedge-Holmes, K. (2016) The University of Glasgow and West of Scotland Local Authority partners: how to engage with MD40 pupils in higher progression schools - SFC Impact for Access Project Report. Glasgow: University of Glasgow.

¹³ Croll, N., Browitt, A., Anderson, M. and Hedge-Holmes, K. (2016) The University of Glasgow and West of Scotland Local Authority partners: how to engage with MD40 pupils in higher progression schools - SFC Impact for Access Project Report. Glasgow: University of Glasgow.

¹⁴ Scottish Government (2016) Introducing the Scottish Government Index of Multiple Deprivation 2016. Edinburgh: Scottish Government.

¹⁵ CoWA (2016) Commission on Widening Access - Technical paper on measures and targets - March 2016. Edinburgh: Scottish Government.

Reliability: Applicants are matched to SIMD20/40 areas based on the home postcodes they declare on their UCAS application forms, which could be verified. Information about an applicant's postcode may be misleading for a small number of applicants such as those who are living in care or who are boarding in a fee-paying school and give the school postcode as their home address, and incorrect for further cases due to mobility and other issues. There is also a substantial number of students in the HESA database every year without 'home' addresses. The 38 measures which make up SIMD are derived from high quality administrative data, with relatively little in the way of missing data.

2.9 Resident in a POLAR quintile 1 or 2 postcode area

The POLAR indicator is readily available to universities via UCAS.

Validity: This indicator relates to residence in an area where the young higher education participation rate is in the bottom 20% or bottom 40% nationally (UK-wide) according to the Participation of Local Areas measure (POLAR). The young higher education participation rate is measured as the proportion of 18 to 19 year olds participating in higher education in a particular census ward. By definition those in areas ranked in the lowest POLAR quintiles are significantly less likely to progress to higher education; currently, the difference in the higher education participation rate between those from bottom quintile and top quintile areas is twenty-five percentage points (UCAS 2015). POLAR quintiles have been shown to be associated with other measures of disadvantage within England, including the Index of Disadvantage Affecting Children (IDACI), social class of parental occupation (NS-SEC), and parental higher education, and school achievement at key stage 4.¹⁶ Census wards contain on average 5,500 eligible voters, but this number is often rather larger in urban areas.¹⁷ Although an analysis by HEFCE found that young participation rates at sub-ward-level (average population size of 1,500) were not substantially different from rates at ward level,¹⁸ it is still the case that wards and sub-wards aggregate information for a large number of people and so the POLAR measure will yield many false negatives and many false positives. Moreover, wards that are more socioeconomically heterogeneous are likely to yield more false

¹⁶ HEFCE (2014) Further information on POLAR3: An analysis of geography, disadvantage and entrants to higher education. Bristol: HEFCE.

¹⁷ See <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/electoral-wards-divisions/index.html>

¹⁸ HEFCE (2014) Further information on POLAR3: An analysis of geography, disadvantage and entrants to higher education. Bristol: HEFCE.

negatives and more false positives than wards which are more socioeconomically homogenous. In addition, false negatives may be more common in Scotland than in other parts of the UK given that participation rates are generally higher in Scotland relative to the UK average. For this reason, POLAR is not used as a method of setting or measuring progress towards widening participation in Scotland; however, it is still the case that Scottish census wards that are in the bottom 20% or 40% of the POLAR distribution are areas with low higher education participation rates in the wider UK context and all the more so in Scotland.

Reliability: Applicants are matched to POLAR quintiles based on the home postcodes they declare on their UCAS application forms, which could be verified. The latest version of this indicator (POLAR3) focuses on young people who were aged 15 at the time of the 2001 census to determine what percentage in each electoral ward entered some form of higher education before the age of 20. Wards are ranked according to their young higher education participation rate and divided into five quintiles each covering twenty percent of all young people. POLAR has been updated twice since its inception in 2005, once in 2007 (POLAR2) and again in 2012 (POLAR3). The data underpinning the latest version of this indicator is taken from the census, cross-referenced against student records maintained by the Higher Education statistics agency. Both data sources are verified administrative records, with relatively little in the way of missing data.

2.10 Resident in an ACORN category 4 or 5 postcode area

Currently only two Scottish universities mention the use of ACORN data in relation to contextualised admissions, but it is used by a number of other universities located elsewhere in the UK. The ACORN indicator is available to universities via purchase from *CACI*.

Validity: This indicator refers to residents in areas characterised as 'Financially Stretched' (ACORN 4) or facing 'Urban Adversity' (ACORN 5) according to the sociodemographic classification system produced by the commercial organisation *CACI*. This indicator classifies UK postcodes into six categories of area (reduced down from 300 smaller categories) indicative of the sociodemographic characteristics, attitudes and lifestyle behaviours typical of those who live there. ACORN category 4, 'Financially Stretched', covers postcodes containing mixed housing types and tenures in which residents typically have incomes much lower than the national average, are more likely to be unemployed and/or in receipt of welfare benefits, and typically are not educated to degree level. ACORN category 5, 'Urban Adversity', covers postcodes containing typically small and often social rented homes on low rise estates or in high rise flats, whose residents are among the most economically and socially deprived

nationally with low incomes, low levels of qualifications, employment mainly in semi-skilled or unskilled occupations, high rates of unemployment and a high likelihood of claiming social security benefits. ACORN categories are statistically associated with school achievement and higher education progression rates. The ACORN indicator would appear to make use of full postcodes which typically relate to a small number of households whose residents may have very similar circumstances. This suggests ACORN will be less prone to false negatives and false positives than other area-level indicators such as SIMD and POLAR where the number of households in each locale is larger and more diverse. However, it is still not a measure of the personal circumstances of specific individual, and its commercial nature means that the precise methodology used to identify ACORN categories is somewhat opaque.

Reliability. Applicants are matched to ACORN 4/5 categories based on the home postcodes they declare on their UCAS application forms, which could be verified. The data underpinning ACORN comes from a mixture of generally high quality administrative, commercial and survey sources, although these are all likely to be subject to missing data to some extent. The ACORN typology is updated annually.

2.11 Resident in a rural area

This indicator not currently available to universities via UCAS, but a postcode look-up file could be made available by the Scottish Education Analytical Services division for schools in Scotland and by equivalent bodies for England, Wales and Northern Ireland.

Validity. Research has highlighted that disadvantaged students living in rural areas are less likely to be picked up by area-level measures of disadvantage such as SIMD or POLAR than similarly disadvantaged students living in more urban locales.¹⁹ Disadvantaged students living in rural areas may not have access to a full range courses at secondary education level,²⁰ and the remoteness of their location may impede progression rates given the often significant distance, time and cost involved in travelling between home and higher education institutions.²¹ However, living in a rural area by itself does not necessarily denote disadvantaged status and so a simple 'rural' indicator will yield a substantial number of false

¹⁹ Skerratt, S. et al (2014) *Rural Scotland in Focus 2014*. Edinburgh: Rural Policy Centre, SRUC, Scotland's Rural College.

²⁰ Lasselie, L. (2016) Barriers to higher education entry – a Scottish rural perspective, *Scottish Educational Review*, 48(1).

²¹ Lasselie, L. and Macpherson, R. A. (2017) Hard to reach communities and a hard to reach university. *Journal of Academic Perspectives*, 17(1)

positives. False negatives will also of course be numerous. More research is needed to identify a valid indicator of rural disadvantage.

Reliability: As with other area-level measures, living in a rural area can be determined from applicants' self-declarations of their home postcodes, which could be cross-referenced against an urban/suburban/rural categorisation of postcodes held by the Scottish Education Analytical Services division for schools in Scotland.

SCHOOL-LEVEL INDICATORS

2.12 Attended a non-selective, non-fee-paying school or college

This indicator is not currently in use by Scottish universities but is included here because much of the evidence used to justify contextualised admissions policies points to research findings showing that students educated in non-selective non-fee-paying secondary educational institutions can perform better at degree level than privately educated students admitted with the same level of school attainment (see Report 4 for a fuller discussion of this evidence). This indicator is available via UCAS.

Validity: This indicator can be thought of as a proxy for relative disadvantage at the individual-level, or as an indicator of a relatively disadvantageous educational environment whether due to a presumed lower quality of teaching or to a presumed more challenging learning environment. Of course, many of those who attend non-selective state schools are in fact not themselves disadvantaged, nor are they necessarily disadvantaged by virtue of their educational environment, and so this indicator will yield a large number false positives. Conversely, a minority of those attending selective non-fee-paying schools (i.e. grammar schools) and selective fee-paying schools (i.e. private schools) are from socioeconomically disadvantage backgrounds, meaning that this indicator is likely to yield at least some false negatives.

Reliability: Information about school(s) attended is entered by applicants on their UCAS forms, is presumably verified by teachers who subsequently provide applicant references, and could be verified by schools and exam boards. Type of school could then be determined from this information referenced against administrative data sources. As such this indicator is likely to be highly reliable for those applying whilst still at school, although there will inevitably be some degree of missing data, for example for those who were home-schooled or educated abroad.

2.13 Attended a school with a high percentage of FSM and/or EMA recipients

Data relating to school-average rates of FSM and EMA receipt is available to universities via UCAS for schools in Scotland and Wales and can be obtained by universities from the Department for Education National Pupil Database for schools in England.

Validity: This indicator is not in use by Scottish universities as a contextual indicator for the general applicant but is included here because it may be a useful school-level proxy of individual-level disadvantage if information about the FSM or EMA status of individual applicants is not available to universities at the point of admissions decision-making. It might also be considered as an indicator of a more challenging educational environment. However, it should be noted that this indicator is only really of value if the same information is not available at the individual level, since area-level measures are inevitably more prone to false negatives and false positives than the same information measured at the level of the individual. In any case, some thought needs to be given as to the threshold at which the percentage of FSM/EMA recipients is deemed high enough to warrant a contextual flag. A high threshold will mean fewer false positives but more false negatives, and *vice versa*.

Reliability: The data for this indicator are based on verified administrative sources of information and so can be considered highly reliable in the sense of being likely to yield few if any false positives. However, since the data refers to receipt of free school meals as opposed to eligibility for receipt, and since not all eligible families provide the information needed to assess their eligibility, some false negatives are likely.

2.14 Attended a school with low average levels of attainment and/or progression to higher education

Data relating to school-average levels of pupil attainment at Highers level and at GCSE and A-level is available to universities via UCAS or directly from SQA. Data on school-average levels of progression to higher education can be obtained by universities from the Scottish Education Analytical Services division for schools in Scotland and from the Department for Education National Pupil Database for schools in England.

Validity: This indicator focuses on schools and colleges in which the average attainment of pupils in Highers (or GCSE and A-level for those educated in England and Wales) is low relative to the national average, or in which rates of progression to higher education are low relative to the national average. It includes but is not necessarily limited to those schools

targeted by the SHEP and ADHP outreach programmes. School-average levels of attainment and school-average rates of progression to higher education are highly correlated. Both can be considered imperfect proxies for socioeconomic disadvantage at the individual level, or as indicators of a more challenging educational environment. Those who attend low-achievement or low-progression schools are, by definition, statistically less likely to gain access to degree-level study, although as with all indicators which refer to aggregates rather than individuals this may not hold in some individual cases. Some thought needs to be given as to the threshold at which the school-average attainment levels or progression rates are deemed low enough to warrant a contextual flag. A low threshold will mean fewer false positives but more false negatives, and *vice versa*. Using this indicator would be perverse if it meant penalising students from schools that are doing very well with disadvantaged students, and favouring instead more advantaged students from poorly performing schools. Equally, it cannot be assumed that all students attending high achievement/progression schools are necessarily socioeconomically advantaged. Research on pupils in schools in the West of Scotland indicates that high progression schools serve a significant number of students from SIMD40 postcodes, and that the gap in attainment and progression rates between SIMD40 and non-SIMD40 students is wider in high progression schools than in low progression schools.²²

Reliability: These data are based on verified administrative sources of information and so are likely to yield few false positives and to have minimal missing data.

3. Key findings and recommendations

3.1 Key findings

Table 1, below, summarises the key findings of our conceptual evaluation of the validity and reliability of a range of potential indicators of contextual disadvantage. The first four columns refer to the likely prevalence of false positives and false negatives in relation to validity and reliability respectively. The final column lists our preliminary recommendations for each indicator.

²² Croll, N., Browitt, A., Anderson, M. and Hedge-Holmes, K. (2016) The University of Glasgow and West of Scotland Local Authority partners: how to engage with MD40 pupils in higher progression schools - SFC Impact for Access Project Report. Glasgow: University of Glasgow.

Table 1. Assessment of the validity, reliability, and overall suitability of potential indicators

	Validity		Reliability		Suitability For use on an EITHER/OR vs an AND/ALSO basis
	False positives	False negatives	False positives	False negatives	
Individual-level					
First-generation in family to gain a degree	Some	Few	Some	Some	Unsuitable
Parents not in professional or equiv. employment	Some	Few	Some	Some	Unsuitable
Has spent time in care	Few	Many	Few	Some	EITHER/OR
Carer for a family member	Few	Many	Few	Some	EITHER/OR
Refugee/asylum seeker	Few	Many	Few	Some	EITHER/OR
Receives free school meals	Few	Some	Few	Some	EITHER/OR
Receives an Education Maintenance Allowance	Few	Some	Few	Some	EITHER/OR
Area-level					
SIMD20/40 postcode	Some	Many	Few	Few	AND/ALSO
POLAR quintiles 1 and 2 postcode	Some	Many	Few	Few	AND/ALSO
ACORN4/5 postcode	Some	Many	Few	Few	AND/ALSO
Rural area	Many	Many	Few	Few	Unsuitable
School-level					
Non-selective non-fee-paying school	Many	Few	Few	Few	Unsuitable
School with a high % pupils receiving FSM/EMA	Some	Many	Few	Few	AND/ALSO
Low attainment/low progression school	Some	Many	Few	Few	AND/ALSO

Our recommendations are as follows:

- The following indicators can be considered **suitable for use alone on an EITHER/OR basis**, because of their likely low incidence of false positives with respect to both validity and reliability: has spent time in care, is a carer for a family member, is a refugee or asylum seeker, received free school meals, or is in receipt of an Education Maintenance Allowance.
- The following indicators are **potentially suitable for use in combination on an AND/ALSO basis** on the grounds that they are likely to yield some (i.e. more than just

a few, but not as much as many) false positives in relation to validity if used singly, but fewer false positives if used in combination: resides in a SIMD20/40 postcode, lives in area classified as quintile 1 or 2 on the POLAR measure, resides in an ACORN 4 or 5 postcode, attends a school with a high rate of pupils in receipt of free school meals and/or an Education Maintenance Allowance (though this particular measure is not recommended for use if the same information is available at the individual level), and attends a school where pupil attainment and/or progression to higher education is low on average.

- Several indicators are ***unsuitable for use alone or in combination***, at least in their current form, due to a likely high incidence of false positives. These indicators include individual-level measures that are not readily administratively verifiable, such as first generation in the family to gain a degree, and parents not in professional-level or equivalent employment; and indicators which define socioeconomic disadvantage too broadly, such as living in a rural area, and attended a non-selective non-fee-paying school..
- Currently, not all of the indicators listed above as suitable for use on an EITHER/OR or AND/ALSO basis are readily available to higher education institutions in verified form. For indicators where the number of applicants concerned is small (e.g. applicant has spent time in care), or where contextual information can be added easily and free of cost (e.g. the non-proprietary postcode based measures SIMD and POLAR), it is feasible for institutions to add these contextual indicators to applicant records on a case by case basis. However, it would be helpful for the sector as a whole if all appropriate indicators were provided directly to institutions in verified form by a third party such as UCAS.
- Further empirical research is needed to estimate the extent to which different indicators are likely to produce false positives and false negatives, and to determine the optimal use of specific indicators on an EITHER/OR and AND/ALSO basis.