Algorithmic Impact Assessment of the predictive system for risk of homelessness developed for the Allegheny County

Eticas Research and Consulting
August, 2020

Research team

Carlos Castillo (Universitat Pompeu Fabra), Mariano Martín Zamorano, Giovanna Jaramillo, Sara Suárez Gonzalo (Eticas Research and Consulting).

Index

1- Introduction 4

1.1 Problem definition 5

1.1.1 Algorithmic development and model 5

1.1.2 How the algorithm is used 7

1.2 Aims of the audit and methodology 10

2- Algorithmic discrimination, fairness and desirability 13

2.1 Algorithmic discrimination 13

2.2 Algorithmic fairness 15

2.3 Algorithmic desirability 17

Desirability “by design” 18

Increasing algorithmic desirability through accountability: 19

3- Homelessness and its determinant factors 21

4- Social setting and homelessness in Allegheny 25

4.1 Socioeconomic situation in Allegheny 25

4.2 Homelessness in Allegheny 26

5- Algorithmic processing in Allegheny public services and bias 30

5.1 Automation of public services in Allegheny County and the algorithm for homelessness 30

5.2 Historical bias and hypotheses about algorithmic discrimination 33

5.2.1 Hypothesis concerning the accuracy of the model 35

5.2.1.1. Internal validation of the model 35

5.2.2 Hypotheses on algorithmic discrimination 36

5.2.2.1 Internal validation for groups 40

6- Analysis of the dataset composition 40

6.1 Age groups within the dataset 40

6.2 Race-based assessment of the dataset 42

6.3 Gender-based assessment of the dataset 45

7. Disparate Impact and Disparate treatment assessment 48

7. 1. Introduction 48

7.2 Analysis of single cases (Family_children==0) 49

7.2.1 DI/DT assessment by race: outcomes Mhinp, ER4 and New Jail 50

7.2.2 DI/DT assessment by gender: outcomes Mhinp, ER4 and New Jail 51

7.2.3 DI/DT assessment by age group: outcomes Mhinp, ER4 and New Jail 52

7.2.4 DI/DT assessment by disability status: outcomes Mhinp, ER4 and New Jail 54

7.2.5 DI/DT assessment by veteran status: outcomes Mhinp, ER4 and New Jail 55

7.2.6 False negative rate (FNR) by group: outcome MH in-patient 56
7.2.7 False negative rate (FNR) by group: outcome ER4 57
7.2.8 False negative rate (FNR) by group: outcome New Jail 57
7.3. Analysis of family cases (Family_children==1) 59
7.3.1 DI/DT assessment by race: outcomes Mhinp, New Jail and ER4 59
7.3.2 DI/DT assessment by gender: outcomes Mhinp, New Jail and ER4 61
7.3.3 DI/DT assessment by age group: outcomes Mhinp, New Jail and ER4 62
7.3.4 DI/DT assessment by disability status: outcomes Mhinp, New Jail and ER4 63
7.3.5 DI/DT assessment by veteran status: outcomes Mhinp, New Jail and ER4 64
7.3.6 False negative rate (FNR) by group: outcome Mhinp 65
7.3.7 False negative rate (FNR) by group: outcome New jail 66
7.3.8 False negative rate (FNR) by group: outcome ER4 67
8- Analysis 68
8.1 Dataset composition, model training and overall accuracy in risk assignation 69
8.1.1 Protected groups within the dataset 69
8.1.2 Intersectional structure of the training data 69
8.1.3 Overall identified accuracy in risk assignation 70
8.2 Bias and potential discrimination within the system 71
The majority (53.2%) belongs to the group between 26-45 years old. 74
8.3 Desirability analysis 79
8.4 Recommendations 82
8.4.2 Algorithmic discrimination 82
8.4.3. Future re-modeling 83
8.4.4 Explainability 84
8.4.5 Desirability 84
References 86

1- Introduction

The Allegheny County Department of Human Services (DHS) has implemented an algorithm aimed at **effectively prioritizing supportive housing for people who are at higher risks of harms associated with homelessness**. It is a tool used to prioritize individuals in need of temporary or permanent housing so that the most disadvantaged can be supported first. Depending on the housing options available at the Allegheny social services, it can be used for placement in housing as well as eligibility for other housing services. Consequently, **the system is oriented towards allocating benefits based on a partially automated risk assessment**.

The system is aimed at addressing the increasing demand for homelessness programs within the County by improving its prioritisation mechanisms. This new Predictive Risk Model will replace an existing actuarial tool used with a similar purpose. In this context, the system should be able to make the existing prioritization process faster and more accurate, by automatically weighing the information provided by the client in the context of the overall population, when assessing his or her level of vulnerability. In order to do this, the system processes both data coming from different databases of the Allegheny County and the information provided by the person calling to the homeless services coordinated hotline. The model allows the DHS to analyze this information to predict the risk of homelessness that the client may experience in the future based on multiple variables.

This document reflects the first part of the assessment of this algorithm conducted by Eticas Research and Consulting. It addresses both the accuracy of the algorithm in regards to its expected outcomes and its potential for discrimination against specific social groups. To address these two levels of analysis, the team in charge of the assessment has followed an iterative process of reconstruction of the model. This "extended model card" analysis includes the recreation of the whole prioritization governance around the algorithm, the analysis of its theoretical basis and the study of the current situation with homelessness in Allegheny, which allows us to infer existing historical bias. After explaining these registers of analysis in the first three sections of this document and shortly introducing our methodology, we will describe the hypotheses of bias and identify the protected groups to be considered in the technical assessment of disparate impact and treatment in section 5. Lastly, section 6 describes the training dataset composition, number 7 explains the results of this quantitative analysis and section 8 reflects the conclusions, which go back to the analysis of the model, also proposing preliminary measures to mitigate and prevent bias.

1.1 Problem definition

1.1.1 Algorithmic development and model

To develop and test the new algorithmic system, the team in charge of its development trained it with historical data corresponding to 5,550 observations from assessments conducted by **link staff of the Allegheny Department of Human Services (DHS)** between January 2016 and March 2017. This information corresponds to calls at the time a person calls the homeless services coordinated hotline. This testing was used to train models on four target outcomes:

I. **MH Inpatient**: At least one inpatient **mental health service** in the 12 months following the call;
II. **Jail Booking**: At least one Allegheny County **Jail booking** in the 12 months following the call;
III. **ER 4+ Visits**: More than four **Emergency Room** (**ER) visits** in the 12 months following the call;
IV. **Substance Use Svcs**: At least one **substance use services** contact in the 12 months following the call.

The above four models were selected based on the **state of the art, information provided by experts in the field and actual data about homelessness**. The Table 1 below, provided by the research team, lists the few instances where the scientific literature has discussed the consequences of homelessness. The final approach was discussed with researchers in two workshops.

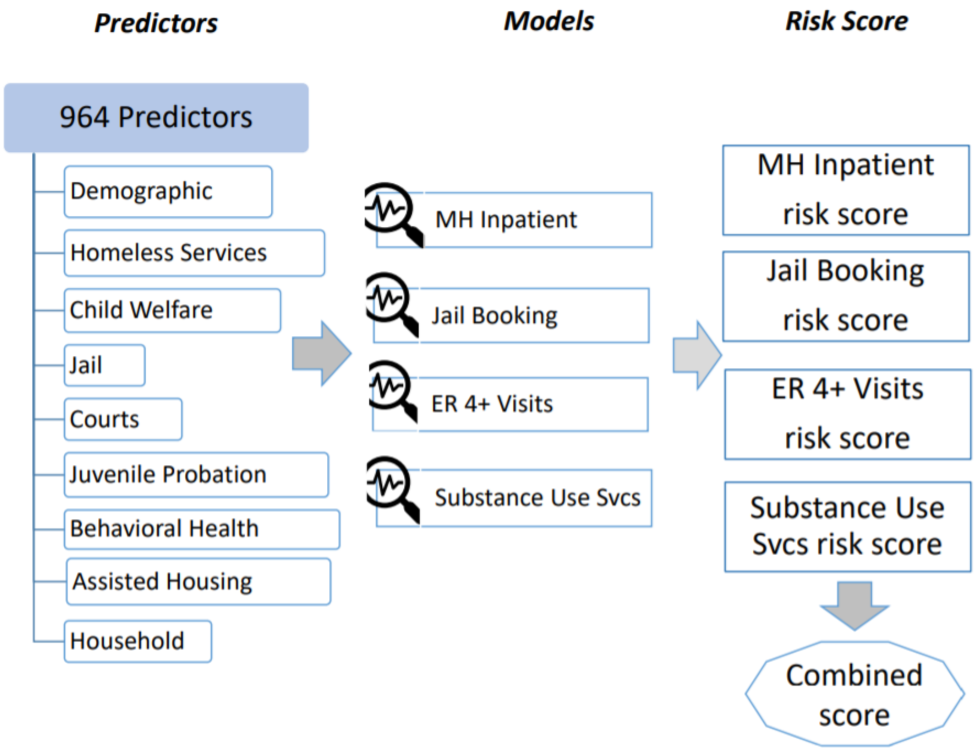
Table 1. Theoretical basis for setting the target outcomes

Outcome / Harm	Homelessness Research Literature
MH Inpatient	(Schütz, 2016) (Bradley, 2018) (Toros, 2018)
ER 4+ Visits	(Cheallaigh, 2017) (Toros, 2018) (Clark, 2019)
Jail	(Weiser,2009) (Mitchell,2017) (Gonzalez,2017) (Toros, 2018)
Substance Use disorder	(Clark, 2019) (Schütz, 2016) (Toros, 2018)

Source: Own elaboration

To train the algorithmic model based on the above four target outcomes, **964 variables ("predictors") from nine domains** were used, which correspond to the categories shown in the following image:

Figure 1. Predictors, models and risk score (initial design)



Source: Kithulgoda, et al. (2019)

The impact of each variable (any of the 964 predictors) is **learned and optimized in the model training phase of Predictive Risk Modelling (PRM)**. Each model has been trained to recognize the optimum mapping between sets of predictors and the corresponding outcome (an actual outcome that indicates the absence or presence of harm); the impact of each variable was determined by a standard machine learning algorithm (Lasso-regularized Logistic Regression[1]). The model was modified in 2019 by removing the outcome "Substance Use Svcs: At least one substance use services contact in the 12 months following the call", since it undermined prediction accuracy level. This audit is therefore conducted on the other three outcomes mentioned above.

The PRM tool is expected to be deployed in 2020.

[1] LASSO regularized Logistic Regression was the machine learning algorithm of choice. Each model was instantiated through the R package *glmnet*. Predicted risk probabilities were grouped to 20 equal–sized bins defined by quantiles.

1.1.2 How the algorithm is used

The risk assessment provides social services **with relevant information about how to classify individuals at risk of experiencing an adverse outcome**. On this basis, a set of the highest score families and individuals are provided with public assistance, including a designated **number of housing units for families and individuals**. If there is a vacancy for a family unit, then this vacancy is filled with the next highest scoring family and the same is done in the case of individuals. There are programs oriented towards families and individuals, and vacancies are filled similarly.

Since the DHS receives more than 10.000 requests of homelessness services annually[2] and there is a limited amount of resources dedicated to them, support seekers need to be classified and prioritized. In fact, according to the Center for Social Data Analytics (2017), at "any time there are around 550 households waiting for services lots but only 50 slots become free each month." In this regard, it is evident that there is an important shortage of resources, visible through the disproportion between the demand for Rapid Re-Housing and Permanent Housing by Allegheny citizens and the offer available for these services in the County. For instance, while **single individuals demanding these programs by July 2019 had reached 542, there were only 30 beds available per month and a subset of 94 people within the waitlist was identified as chronically homeless**.

There is a designated **staff person (link staff) whose full-time job is to work with clients and service providers to ensure that both parties are ready for the referral**. This staff receives calls from residents facing housing instability or homelessness and triages them for programs and support services. This person is in contact with the client, shelter staff and others. His/hers role is to talk with the clients about their needs and experiences, and about their history and current situation. This helps the staff to evaluate whether the score the algorithm gives to a client is adequate or if it is not aligned with what he/she has heard from the client. With this information in mind, the staff can also evaluate whether it is necessary to go through another question-answer process that will assign an alternative risk score to this client.

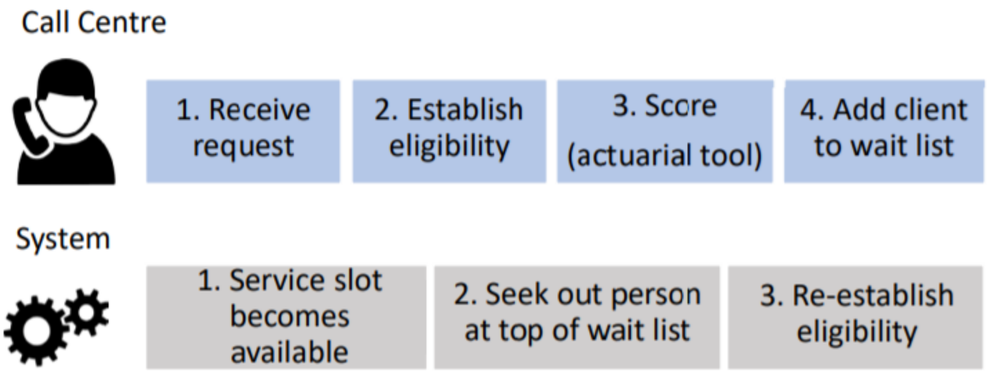
This staff member attempts to contact the client a minimum of three different times throughout the day and make contact with other service providers which are engaged with the client at the same time (i.e. street outreach staff, shelter caseworkers, etc.). If the client cannot be contacted after three attempts, then the next person on the list becomes prioritized. However, the previous person is not removed from the priority list and becomes eligible to fill the next vacancy.

---

[2] Including homeless support, homeless prevention, street outreach, emergency shelter, bridge housing, permanent supportive housing and rapid rehousing.

There is a second, separated, person of the staff working on the top of the list to assess how resources have to be allocated. In this second step, taking into account other information handled by the County, as well as the availability of resources, the staff prioritizes and contact the persons on the top of the list, according to what she/he can handle. This implies that the exact order of prioritization in the list is not followed, but rather the order that can be managed by the staff.

Figure 2. Prioritization process



Source: Kithulgoda, et al. (2019)

Along these lines, the **main aim of the algorithmic system is prioritization**, which is conducted partially based on the user's answers to a series of questions to be provided at the call centre. The score resulting from the processing of these data, together with data already held by the DHS, is used as part of the decision-making process. In particular, the risk score associated with each client is used by DHS administrative and social workers to conduct the final assessment and evaluate eligibility concerning the selection and provision of services. The system, then, is not designed to decide on the recipient of the service automatically, but it assists in decision making by providing human staff with a risk score which helps to rank clients for service.

According to the DHS, two main objectives for the implementation of this predictive system are: 1. to prevent the staff from asking very sensitive questions to the persons that are calling (who may experience an unpleasant situation), and 2. to use the robust data warehouse the DHS has, so as to be more accurate and to be faster in both the prioritization process and the allocation of resources.

The DHS informs that the human staff making these decisions has gone through training provided by the Department of Human Services and other trainings. As part of these training activities, the staff received information about how their duties will change in relation to the previous methods used to prioritize clients, how the tool is working and how it will change, and how the clients will be scored based on case reviews. Moreover, the staff has gone through different case reviews so as to become comfortable with the system and to understand how the people will be scored. More formal training will be done when they are actually interacting with the system that has not been conducted yet.

The results of this process in terms of service provision are translated into the classification of users as homeless, chronic homeless or not homeless, following the criteria below:

Table 2. Allegheny and DHS classification

Homeless people
Street Outreach: Persons living on the streets or other places not meant for human habitation are engaged by street outreach workers to provide basic needs (food/water/medical care) as well as to connect them to housing services and other supportive services (behavioral health, etc.).
Emergency Shelter: A facility with overnight sleeping accommodations, the primary purpose of which is to provide temporary shelter.
Winter Shelter: This emergency shelter provides sleeping accommodations between 7 p.m. and 7 a.m.
Transitional Housing: A facility designed to provide housing and appropriate supportive services to homeless people to facilitate movement to independent living within a reasonable amount of time, usually 24 months.
Safe Haven: A form of supportive housing that serves hard-to-reach **homeless people with severe mental illnesses** who come primarily from the streets and have been unwilling to participate in housing services. These types of shelters currently serve eligible veterans.
Rapid Re-Housing (RRH): Programs that assist individuals or families who are experiencing homelessness to move as quickly as possible into permanent housing and achieve stability in that housing through a combination of rental assistance, housing search and supportive services.
People not considered to be homeless
Permanent Supportive Housing (PSH): Combines housing with more intensive services for those with **one or more chronic disabling conditions** and does not have a limit on length of stay as long as the tenant pays their portion of the rent and follows the rules of their lease.

Source: Own elaboration based on DHS (2018).

As explained by Kithulgoda, et al. (2019: 2), “All clients who are homeless are eligible for shelter services, however housing programs of longer duration, such as rapid rehousing, transitional and permanent supportive housing, require clients to meet a higher threshold”. Therefore, the **unexpected externalities of the system mostly relate to the incorrect classification of individuals at high risk of homelessness. This may affect them by reducing their social protection or by not assigning them to an appropriate social program**. In this regard, whether the system is allocating permanent supportive housing resources to the "right" people should be regularly contrasted against the established definitions about each service and its goals. Moreover, such **misclassification might be based on protected attributes corresponding to already disadvantaged groups, such as disabled or Black/African American people, leading to discrimination**. In this regard, false negatives- meaning consistently assigning a low risk of homelessness to individuals needing assistance belonging to certain groups-, could lead to not assisting to already disadvantaged groups. These are protected groups such as black people or economically extremely disadvantaged people.

1.2 Aims of the audit and methodology

Our methodology for this analysis is mainly oriented towards determining both the accuracy of the algorithm in regards to its expected goals and identifying potential discrimination derived from its decisions. To achieve this, before the final phase of the analysis, we followed three main steps. Firstly, we reconstructed the model design and the theoretical and methodological basis established for the algorithmic model. In this context, we reviewed the literature on homelessness to have a basis for valuing the desirability of the model. The developing team provided valuable inputs and answers to our information requests to complete this process.

Secondly, we developed a statistical analysis of the situation with homelessness in the USA and particularly in Allegheny. This is aimed at framing the main factual causes leading to homelessness in the social context where the system is implemented, as well as constructing hypotheses about ways of accurately measuring the risk of homelessness under these social circumstances. The above activities were also oriented towards establishing a set of hypotheses about algorithmic fairness, accuracy, and discrimination. Thirdly, on this basis, we followed the following steps: a) define an assignment of elements in the data to groups, b) define protected groups, c) determine a set of metrics aimed at measuring bias, and, d) measure and compare across groups.

a) The first step simply sorts the data items into groups, which can be overlapping ("soft" assignment) or non-overlapping ("hard" assignment). In most cases, the data items would correspond to people, and hence, the groups will be done on individual characteristics. Any characteristic of individuals can be used to create such groups, but particular attention is placed in *protected* characteristics. Protected characteristics correspond to attributes of people that anti-discrimination law (including international law) specifically mentions.[3] These groupings are created in

[3] Such as the Civil Rights Act of 1964 in the US, or Article 21 of the EU Charter of Fundamental Rights.

the data to evaluate the extent to which an algorithm may treat or affect a group differently from another. In this regard, we defined algorithmic discrimination as the **systematic production of disadvantageous outcomes against socially salient groups, particularly disadvantaged groups**.

b) The second step consisted of defining which of the groups that have been defined will be protected, meaning that they must not be further disadvantaged by the application of the algorithm and that the impact of the algorithms on them will be specially monitored. In some cases, protected groups are categories that are legally protected (e.g., people with disabilities). In other cases, the definition of what constitutes a protected group is related to a commitment that may not be legally binding, such as an intention to increase the participation of women or minorities who might be underrepresented in certain positions. A further definition of protected group might originate in the purpose of an application and hence in the algorithm's desirability; for instance, if the intention of a certain algorithm is to increase the protection of children of a certain age in an algorithm to screen calls reporting domestic abuse (Chouldechova et al, 2018), then the children of that age constitute a protected group for the purpose of algorithmic bias analysis.

c) The third step determined the set of metrics to be used. In general, these metrics quantify the extent to which an algorithm **treats** people differently (disparate treatment) and the extent to which an algorithm has a different **impact** on different people (disparate impact). Metrics applied to this case were oriented towards assessing the proportion of people that receive negative/unfavorable outcomes across groups, which is a measure of **risk**, which should be equal if we want to claim that the algorithm has equalized risks for protected and non-protected populations (i.e., the **Independence** criterion of group fairness described by, e.g., Barocas and Hardt 2017).

d) In the fourth step, after selecting the above metrics, the data were analyzed to obtain values for these measurements. The **computation of metrics can be done using a combination of existing libraries for algorithmic fairness analysis**, including Aequitas[4] and IBM/AIF360[5]. Their usage supports a process having the steps that we have described above. After the measures and confidence intervals were computed, any disparity was noted, analyzed, and reported. When there is structure in disparities, such as patterns obtained by mining the data annotated with cases in which there is disparity, this structure may constitute a potentially discriminatory practice.

Given that this audit is not only aimed at analysing the efficiency of the system in itself, but also its **explainability** and **desirability**, also from a **qualitative perspective**, it is very relevant

[4] http://aequitas.dssg.io/.
[5] https://github.com/IBM/AIF360.

to have information about the context in which the system has to be implemented, how will it be integrated within the procedures of the County, and how will the workers interact with it. Having this objective in mind, the audit process involves a necessary exchange of information with the County and the system development team, primarily related to the real use of the system (in context) and its **role in the decision-making process**. The idea of the audit at this point is to understand the type of human-machine interaction within the process of decision-making, and to assess whether the **workers** who interact with the predictive system have an adequate relationship with it. That is to say, that "regardless" of the results of the system, they continue to apply their **professional criteria**.

In the final stage of the process, short interviews/questionnaires with Allegheny County workers that interact with the system are expected to be conducted, so as to have some information regarding their levels of **satisfaction**, **training** and **confidence** in the system.
Besides several exchanges of information with the developing team and the Allegheny County concerning different socio-technical aspects of the model, **a qualitative interview with Andy Halfhill** (Manager of Homeless/Housing Analytics, Office of Analytics, Technology and Planning of the Allegheny County Department of Human Services) was conducted to contrast the human factors around the prioritization process. Mr Halfhill is responsible for the system implementation in Allegheny County. Variables addressed during the interview included: a) the programmes for which the algorithm will be used and the overall policy behind technological adoption, b) the expected level of automation of decision-making, c) type of workers interacting with the system and their training and, lastly, d) their surveyed level of satisfaction with the current prototype.

2- Algorithmic discrimination, fairness and desirability

2.1 Algorithmic discrimination

There are different approaches to defining algorithmic bias. First, we have those definitions that stress the **unexpected character of algorithmic processing outcomes according to its predefined aims**. For instance, for Chen et al. (2019) "*algorithmic bias occurs when an AI model, trained on a given data set, produces results that may be completely unintended by the model creators.*" Second, we find those perspectives stressing that **the incompatibility between the outcomes of algorithmic processing and its expected goals is greatly defined by contextual social factors beyond the creators' expectations,** determining what can be considered as fair or ethical. Akthar et al. (2019) point out, along these lines, that algorithmic bias happens when "*the data used to develop and refine algorithms reflect implicit values of the society in ways that are judged as irrational or unfavorable*." Baeza-Yates (2018), in contrast, separates input data bias from algorithmic bias stating: "*algorithmic bias is added by the algorithm itself and not present in the input data.*" Beyond these conceptual perspectives, the literature increasingly recognizes the **importance of social or cultural factors and the prominent role of technological developers in detecting and mitigating algorithmic bias**. The importance of **implicit values** in data collection, selection and use in this process has been stressed by a number of authors (Binns, 2018; Nissenbaum, 2001). This is one of the reasons why research has put particular emphasis on the ethical issues related to machine learning systems (Bolukbasi et al., 2016).

Bias can be approached in different ways also concerning the characteristics of the systems involved. On the one hand, we have **batch-operation systems based on predictive models**, which can be more easily analyzed through existing metrics. On the other hand, we have interactive systems, **where the interaction with users is more complex**, which require contextual factors to be taken into account in order to establish an accurate definition of fairness (Holstein, 2019). Nevertheless, in both cases, the interaction between algorithms in operation and the ethical values involved in algorithmic decision-making should be considered as the main analytical factor. As explained in this section, this requires going beyond mathematical correctness to ensure ethical compliance.

In order to frame algorithmic bias, we should first **distinguish between different forms of discrimination**. Following definitions by Lippert-Rasmussen (2013), generic discrimination occurs when someone treats a person A worse than s/he would treat another person B because A has some attribute that B does not have. ***Group discrimination*** **happens when such attribute consists of simply belonging to a socially salient group**, i.e., a group in which membership "is important to the structure of social interactions across a wide range of social contexts" (Lippert-Rasmussen, 2013:30), and requires animosity against this group, or the belief that people in this group are inferior, or the belief that they should not intermingle with others. In order to be considered discriminatory, bias should involve one or more of the so-called protected groups, which correspond to the following protected attributes, which is based on the attributes included in the US Civil Rights Act of 1991 and the Equality Act of the UK 2010, Section 4, and the European Charter of Human Rights. It should be noted that this is not an exhaustive list, since it may be adapted or modified, depending on the context[6]:

Table 3. Protected groups and attributes

Protected groups (non-exhaustive)	Protected attributes
Children and Elderly	Age
Disable people (physical and mental)	Disability
Women and Transsexual	Gender and Gender reassignment
Multiple social groups (e.g. African American, etc.)	Race, color, ethnicity
Pregnant	Pregnancy
Muslim, Jews	Religion or belief
Gay people, lesbian people, etc...	Sexual orientation
Low income people	Property

Source: Own elaboration.

Following the above rationale, a more precise definition of algorithmic bias -or algorithmic discrimination- **involves the systematic production of disadvantageous outcomes against socially salient groups, particularly disadvantaged groups**, for instance, manifested as consistently lower risk scores for individuals in a group that does not actually have lower risk than another. This bias is embedded in the mathematical properties of an algorithm[7].

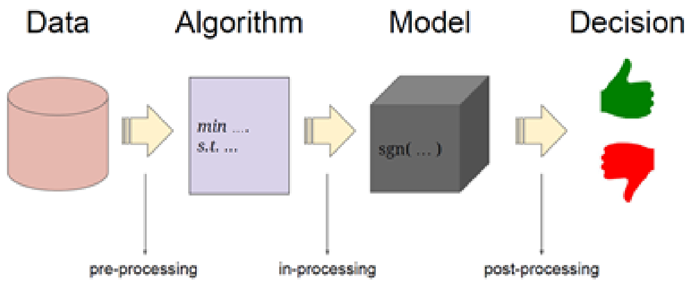
Algorithmic bias has been divided into two different types, depending on the stage of the machine learning process at which it happens (Danks and London, 2017). Firstly, algorithmic bias, as well as biased models, can **be biased due to the collection and use of biased training data when training or modeling algorithms during the initial stages of**

[6] Disadvantaged groups can be defined, e.g., in relation to the attributes mentioned in the Civil Rights Act of 1991 in the US: "race, color, religion, sex, or national origin" or Article 21 (Non- discrimination), of the EU Charter of Fundamental Rights: "sex (and gender), race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation." These protected groups are defined as individuals and groups sharing one or more of the 'protected characteristics'.

[7] We remark that the above definitions are different from standard definitions of statistical bias, which involve distortions of a statistic resulting from biased samples or estimators of which the calculation is not correct in relation to the right or expected value of a parameter (Turney, 1996), and hence statistical bias cannot (always) be an adequate criteria of algorithmic fairness.

development (Cowgill, 2019). Secondly, post-algorithmic or processing bias relates to the **modeling of the system caused by its interactions with users**. The so-called *disparate treatment* of subgroups based on an apparently reasonable basis, therefore, leads to discrimination (Barocas and Selbst, 2016). In this case, the user's interpretation of the outcome of algorithmic processing is key for assessing how fair it is (Baeza-Yates, 2018). The different phases during which algorithmic bias may come about, which are the same phases in which algorithmic bias can be mitigated, are summarized in the image below.

Figure 3. Phases in which algorithmic bias can be mitigated



Source: Hajian, S., Bonchi, F., and Castillo, C. (2016).

2.2 Algorithmic fairness

Consistently with the above definitions, lack of algorithmic fairness could be broadly defined as any case "where AI/ML systems perform differently for different groups in ways that may be considered undesirable" (Holstein, 2019: 3). Even though quantitative methods to capture and measure disparate treatment/impact over disadvantaged groups have been developed, these techniques are not able to encompass the philosophical debate about **which** groups can be considered as disadvantaged and **what** can be considered as differential treatment in a particular socio-cultural setting. In fact, the literature has shown the usual incompatibility between statistical models of fairness and the interpretations made by users or citizens (Kyung Lee, 2018). This debate manifests itself in **multiple definitions of fairness**, which makes it challenging to reach a single accepted definition to be used by scientists and engineers (Narayanan, 2018). In fact, Narayanan (Ibid.) has identified 21 definitions of fairness.

One of the most important definitions concerns **group fairness**, which involves that the algorithmic system in place should not treat specific social groups in an unfair manner. Among group fairness measures we can highlight the three basic ones described by Barocas and Hardt (2017): Independence (also known as Demographic Parity or Statistical Parity), Separation (also known as Equalized Odds or avoidance of Disparate Mistreatment), and Sufficiency (or Calibration), which are three of the most commonly used in the literature. **Independence** means that the probability of assigning an outcome is independent of the protected attribute (e.g., that the probabilities of obtaining a high mental health - MH - risk score is similar across races). **Separation** means that the probability of assigning an outcome is independent of the protected attribute given the actual outcome (e.g., that considering individuals that actually received mental health inpatient service, the MH risk score they received during the evaluation was similar across races). **Sufficiency** means that the outcome assigned by an algorithm does not need to be combined with protected attributes to obtain a prediction (e.g., that a given MH risk score translates to similar probabilities of receiving mental health inpatient services, independently of race).

Some of these metrics of group fairness can be incompatible with each other. For instance, when analyzing systems to predict recidivism, Chouldechova (2018) revealed that an instrument that satisfies predictive parity cannot have the same false positive and negative rates across groups when the recidivism prevalence differs across such groups. Dwork and Ilvento (2018) have comprehensively discussed this issue.

Furthermore, as indicated by Heidari et al. (2018:3) “statistical notions of fairness fail to guarantee fairness at the individual level”. Indeed, a different notion of fairness is **individual fairness,** which was first established by Dwork et al. (2012). For a system to be fair from an individual standpoint, two individuals who are similar in terms of the algorithm’s aims and model must receive similar outcomes. Therefore, this model imposes restrictions on the treatment for each pair of individuals (Kim et al., 2018). However, as pointed out by Speicher et al. (2018), these metrics fail to factor in broader contextual factors, such as differences in previous activities or capitals held by each individual or other social “frames” influencing the outcomes of algorithmic processing. Moreover, according to Speicher al. (2018), there are not efficient computational mechanisms for integrating these sorts of conceptual approaches.

There is an ongoing debate about the development of fairness metrics adapted to different types of algorithms and systems. Moreover, such metrics may be able to overcome existing tensions and trade-offs between different measures of group fairness, as well as between-group fairness and individual fairness. In addition, there is a tricky relationship between fairness and utility, since in some cases functionality can be harmed in order to improve the system’s performance in terms of fairness (Narayanan, 2018). In fact, any methodological approach adopted for the purposes of assessing bias should combine the analysis of specific factors determining fairness and the aims of algorithmic processing. The social context in which the system operates should be taken into account, from both quantitative and qualitative standpoints. Along these lines, as proposed by Castillo (2019), the methods used to rank items such as people, groups, business, or similar, should be able to reach fairness in terms of:

1. A sufficient presence of elements of the protected group

- Absence of statistical (group) discrimination
    - Prevent allocative harms to a group
2. A consistent treatment of elements of both groups
    - Absence of individual discrimination
3. A proper representation of disadvantaged groups
    - Prevent representational harms to a group

The above technical efforts to frame and identify fairness should always be put in a broader theoretical perspective and analyzed case by case. Along these lines, **possible legitimate grounds for discrimination should, therefore, be considered, since in many cases equal treatment is not an expected outcome of algorithmic processing** (Binns, 2018; Dwork et al., 2012). Following the principle of equality of treatment, similar situations should be treated equitably in order to avoid discriminatory treatment when systems make automatic decisions. However, defining which groups or situations must be treated as equal under an ethical criterion requires a contextual-based analysis that properly ponders the dominant ethical grounds in a certain society against the universal values that want to be fostered. In other words, this means considering which might be the impact of the algorithm outcomes on the existing power relations. Moreover, after conducting this relational analysis, a cost-benefit analysis of the trade-offs between the measures is needed to prevent a certain form of discrimination and the functionalities to be achieved by the system must be conducted.

2.3 Algorithmic desirability

Desirability draws on tools from social and policy analysis, with an emphasis on the real implementation and integration of the system. It comprises questions concerning the very need of the proposed solution in a specific socioeconomic context, considering its current conditions, its possible future impact and its alternatives. Hence, the desirability of an algorithmic system will be strictly dependent on both its **capacity to follow the legal requirements** in a specific context, which includes an "operational" definition of proportionality and its **capacity to anticipate and mitigate undesired social impacts**.

Algorithmic design is primarily valued by its capacity to transform input data into "usable" and "desired" outputs (Gillespie 2012). However, valuing algorithmic capabilities exclusively on the basis of its efficiency for accomplishing a defined purpose can limit our understanding of desirability in the context of the information society, where algorithmic decisions based on Big Data have become omnipresent. The analysis of algorithmic desirability should instead go beyond its framed goals and be based on both its expected and potential social impacts.

The social impact of an unfair algorithmic decision process could involve reproducing or deepening socioeconomic inequality (for instance, a bad algorithmic score for bank credits to immigrants, who use to base at the bottom up the pyramid, or selection processes for schools or jobs) stigmatization of social groups and racial discrimination (for instance, conviction rates due to the use of biased information on suspects, Angwin, et al., 2016) and other decisions that affect the right to integrity.

Furthermore, public institutions and governments would significantly improve the efficiency of their policies, which translates into either more money to allot to alternative uses or into tax cuts, both of which can be sold to the electorate in order to gain votes (political motivations) and/or to increase public trust in democratic institutions and processes (social cohesion and stability motivations). For example, if the judicial system is more accurate when it comes to judge convict’s likelihood to relinquish, this can come as a result of an improvement of the algorithm’s equitableness towards minorities or protected collectives. This would allow for a more effective allocation of resources. At the same time, we could avoid the so called ‘’feedback loops’’, a concept coined by Cathy O’Neil (2016) which points out to the potential negative spirals that the use of algorithms can cause in people’s lives. For instance, if a member of an ethnic minority is classified as less likely to pay back and, subsequently, is not granted a loan, this can lead to the commission of a crime, to the inability to find a job and so on. This can mean the loss of a potentially functional member of society and, additionally, higher public spending related to the prosecution and the incarceration of the individual.[8]

Desirability “by design”

It is likely that the background of those creating algorithms has something to do with the inequitable outcomes produced by them (O’Neil, 2016). In the end, data reflects how people operate in society and neither data nor algorithmic system’s assumptions always capture or predict social complex behaviours or structures. As a consequence, algorithms might tend to reward people who conform to a given set of standards, while progressively excluding the rest, mainly the poor and uneducated, but also women (especially those who are single parents), people of colour (any shade of brown or black), ‘old’ people (those over fifty), immigrants, Muslims and non-English speakers. Those are all collectives that will likely be negatively affected by the logic embedded in algorithms which have been built by individuals belonging to certain portions of society.

Moreover, algorithms are increasingly complementing or replacing human judgment or, at the least, decisively informing the decision making process. Algorithms can in fact silently steer user’s choices or actions towards certain options without them being aware of it. This raises serious questions about the silent power of algorithms which are often designed not to improve users’ online experience, but rather to boost click through rates, page views or

[8] These are direct costs, but an incarcerated citizen represents not only a source of spending for the government, but also the loss of potential revenue in the form of taxes and various contributions that the individual could offer to society.

time spent online. In such a way, algorithms do not only reduce human autonomy in decision making, but they do so moved by economic motivations.

On a different note, one might also seek to understand to what extent algorithms are actually better than humans at making decisions. On this matter the USA cases about predicting recidivism could probably help to shed some light on the issue. For instance, recidivism predicting tools, such as COMPAS, are not more accurate or fairer than predictions made by humans with little or no criminal justice expertise. Moreover, they also show that a simple linear predictor equipped with only two features is nearly as accurate as COMPAS, with its 137 survey questions.

Moreover, the debates about predicting recidivism brought other issues to the surface, regarding the purpose for which algorithms are used. Barabas et al. (2018) argue that the tools for risk assessment, currently used for predicting recidivism, should instead be used for helping in diagnosis and intervention. As they show, the regression analysis and machine learning methods currently deployed in risk assessment tools work well in individuating correlations, but do not provide any help for understanding the causal relations among crimes and the reasons behind them. Their claim is that this predictive capability should be used to design appropriate interventions that are effective at reducing the risk of recidivism. In the case of the risk of, e.g., requiring mental health services or emergency room visits, a similar claim could be made.

Increasing algorithmic desirability through accountability:

One of the **main pillars of social desirability is accountability**, including the public explanation of the values embedded onto algorithms which lead to a given outcome. Nevertheless, as it has already been explained by the literature, the complex nature of algorithms can potentially limit this process (Kroll et al, 2017).

The essential variables hindering the accountability process are briefly exposed down below:

- **Opacity:** In some cases, for reasons of business secret the operation of an algorithm might not be revealed, or the algorithm may be available only in machine-readable and not in human-readable or source code form. In other cases, even released source code might be intentionally obfuscated (Pasquale, 2015).
- **Technical complexity**: Even if the algorithm is expressed in human-readable form and no intentional obfuscation is done, source code in a programming language is usually not intelligible or accessible for most of the population. Even an expert who is familiar with a programming language may struggle to comprehend a complex algorithm. This has a direct effect on questions related to desirability, as well as on matters having to do with democratic representation.
- **Algorithmic unpredictability**: All kinds of algorithms - as well as ensemble algorithmic

systems - have been proven to work with a certain level of unpredictability, particularly when processing Big Data. There is no algorithmic procedure to determine if a given non-trivial program will function correctly or even stop on a given input (a problem technically known as the undecidability of the halting problem). Moreover, the predictive or descriptive character of the algorithmic design has a clear influence on the magnitude of this phenomenon; since randomization techniques applied to decision making processes can determine the predictability of the various outcomes.

- **Contextual complexity:** Algorithmic processing always takes place by factoring in a complex set of social variables and values corresponding to the processed data and its correlations (Gillespie 2012). Different technical aspects define the scope of a set of data to be processed by a specific algorithm, such as the quality of the collection and management processes, including its "adaptability" to algorithmic design, and the exclusion of part of the data due to algorithmic equations. Nevertheless, there are also social aspects that determine the amount and characteristics of the data which is to be processed, such as the preeminence of a social group over another within the sample of the population which is to be analyzed. In this regard, it should be noted that algorithms make decisions regarding individuals or groups which are based on whatever data -or previous patterns- "suggest" them, rather than based on the actual behaviour of the individuals" (Anton Vedder and Laurens Naudts, 2017: 4).

The four above explained aspects limiting the potential for accountability of algorithmic decision making are therefore essential aspects which will have to be factored in when carrying out the desirability assessment. They also evidence limitations that will inevitably show up when trying to audit the functioning of algorithms. A **proper auditing of an algorithmic decision making process should not only take into account the technical layout of the algorithm, but also consider the way that the algorithm is actually going to play out, and how that could potentially affect human rights and citizen rights**. Obviously, this analysis should focus on currently disadvantaged, "protected groups" within the dataset to be processed by the algorithm, since they are often the most likely to be discriminated against.

3- Homelessness and its determinant factors

One of the key factors defining the desirability, accuracy and fairness of the algorithm under analysis is the capacity of the four models explained above to **effectively measure the risk of harms associated with homelessness in the County**. However, the efficiency of the selected proxies in reflecting the probabilities of each model or situation to occur can be conditioned by inadequate definitions behind the model as well as by many historical or sociocultural factors. In this section, we offer a brief introduction to the phenomenon of homelessness, which is not aimed at refuting the theoretical basis behind the established model. Instead, it seeks to provide some relevant and complementary elements about the factors leading to homelessness that can be relevant for our study.

Homelessness has been defined as **lacking a "fixed, regular and adequate nighttime residence"** (National Coalition for the Homeless, 2002). However, the literature and policies in the field have framed homelessness going beyond the actual lack of "conventional home" and including aspects such as "a lack of social connectedness; social and family supports and networks" (Community Services Committee, 2007). In this line, it has been emphasized that homelessness is a multidimensional phenomenon with several determinants. According to the Department of Housing and Urban Development of the USA (HUD) (Allegheny County, 2019):

> "a person or family is literally homeless if they fall in one of the following categories: (1) lacking a fixed, regular and adequate nighttime residence, which includes a place not meant for human habitation or a shelter; (2) will imminently lose their primary nighttime residence within 14 days; (3) is an unaccompanied youth under the age of 25 or a family with children that has experienced persistent instability; or (4) is fleeing domestic violence and has no other residence."

Most homeless persons and families can be located because they use services such as shelters and health clinics, but there is a "hidden" homeless population that includes all the people (mainly young adults) that live (or couch-surf) with friends or relatives and those who live in substandard facilities like abandoned houses or cars (Zerger et. al, 2008: 826). According to the National Health Care for the Homeless Council (2017), individuals or families that have moved more than two times in less than two months or live in single room occupancies should be also considered homeless.

Causes of homelessness are very complex to distinguish: for some researchers (Understanding Homelessness, 2019) it is a conflict between systemic and individual causes; for others (Speak, 2019) the drivers are mainly systemic. Among the factors that may lead to an individual being homeless in developed countries, **poor health, disability, drug and**

alcohol abuse, unemployment, low income, and financial problems have been underlined (Community Services Committee, 2007). Two registers grouping these factors have been stressed. On the one hand, many studies point out that the **lack of financial resources for supporting accommodation** is a relevant aspect to be considered (Steinhaus, 2004). On the other hand, **social capital**, namely the social networks of the individual, is also defined as a major element conditioning homelessness (Bartošovič, 2016). According to Narendorf (2017), disrupted social networks, combined with challenging behaviors and fragile family systems, should be considered at the heart of housing instability and homelessness, especially for young adults. This is in line with the so-called "disaffiliation" tradition in the analysis of homelessness (Wallace, 1965). The causes of homeless have been classified in many ways. Bartošovič (2016), classifies them as objective or structural (including the employment or housing policy, social treatment of excluded groups, etc.), subjective or agential (including the loss of jobs, properties, etc.) and "personal", where health problems or addictions have been grouped.

In that same note, Martijn and Sharpe (2006) have identified five different pathways to homelessness, caused by interaction between factors such as **psychological problems, abuse of drugs and alcohol, trauma, family problems, and crime.** This approach in line with Craig and Hodson (1998), who suggested that the best way to understand homelessness among the young population, is to identify pathways, instead of isolated factors. They found that **trauma and family problems were common among young adults before they became homeless.** Wong et al. (2016), have found that a majority of young adults have suffered or witnessed adverse childhood events like domestic violence or sexual violence before homelessness. Different studies on this field (Bender et al., 2014; Ferguson, 2009; Keeshin and Campbell, 2011; Rosenthal, Mallet and Myers, 2006) assert having experienced traumatic events such as physical or sexual abuse in youth are related to a higher risk of homeless.

Those traumatic events, as well as the victimization caused by homeless, have relevant psychological effects. Martijn and Sharpe (2006) found out that 25% of their interviewees with psychological problems developed them after a traumatic event occurred before becoming homeless. Narendorf (2017) carried out a qualitative analysis on 54 young adults in southwest urban areas in the United States and compiled stories on how young men and women describe how **mental health problems started or increased after the harshness of homelessness. Among those psychopathologies are PTSD, anxiety, depressive disorders or self-injuries.** Even though many young men and women seek help for these illnesses, most of them do not use the services available. According to Solorio et al. (2006), only 32% out of 688 young adults surveyed in Los Angeles received treatment. Among the barriers mentioned by Solorio et al. (2006) and other researchers are that mental health services are

not well known or are difficult to access, the distrust of youngsters towards the service providers and the stigma of mental illnesses and the fear of discrimination.
In Martijn and Sharpe's (2006) study, **drug and alcohol abuse was preceded most of the time by a traumatic event and/or by a psychological problem**. This means that in their sample it was more common for young men and women with psychopathologies to develop a drug or alcohol addiction. The experience of transiency is also linked with substance abuse, Bender et al (2014) found out that young adults considered those intercity moves after leaving home for the first time as the ones that disrupted their social networks, exposed them to the unknown and led them to start consuming different substances. Other studies (Bear, Ginzler and Peterson, 2003; Green, Ennet and Ringwalt, 1997) have found that the use and abuse of alcohol and drugs increase with the duration of homelessness and the risk that it implies. In the same note, Zerger (2008) argues that since different drugs have different physiological effects they could be associated with different circumstances and risky behaviours.

There is a propensity to associate homelessness to crime, but according to Martijn and Sharpe (2006), just **one-third of their sample turned to crime to solve their most basic needs or their drug or alcohol addiction**. This means that most young adults that admitted to having committed some type of crime were either traumatized before, suffered a mental illness and/or had an addiction to drugs or alcohol. According to Fischer, Shinn, Shrout and Tsemberis (2008), the likelihood of committing crimes changes as individuals' cycle between episodes of homelessness and housing and as mental illness symptoms become more and less severe. Fischer, Shinn, Shrout and Tsemberis, stress that the most common offenses committed by homeless persons are public disturbances such as sleeping on the parks and urinating on public spaces, or petty crimes like shoplifting, vandalism or breaking into stores. However, some **actions are not considered as crimes but are being criminalized such as camping, sleeping and begging in public or living in vehicles**. In this regard, the National Law Center of Homelessness and Poverty has analyzed laws in 187 US cities since 2009 and has identified that those kinds of city ordinances have multiplied in the last ten years with excesses such as laws that prevent people experiencing homelessness from resting in public spaces.

All this shows that **homelessness is a phenomenon that has many implications in terms of how it can lead to increasing and reproducing social exclusion**. Therefore, the chain of factors leading to this precarious situation must be considered in our analysis. Along these lines, the impact of homelessness on the future development of children and youth has been particularly addressed (Riden, 2011:6). For instance, homeless adults who first became homeless before 50 have shown to have: "more adverse life experiences (i.e., mental health and substance use problems, imprisonment) and lower attainment of adult milestones (i.e., marriage, full-time employment) compared to individuals with later onset" (Brown et al.

2016). Moreover, Brown et al. (2016) have analysed the variations in the incidence of different factors according to age groups of adults from 18 to 25 and from 26 to 49, showing that **economic and health aspects are prevalent in the older group**. **Drug/alcohol abuse and imprisonment at a young age** have an important influence in the case of the younger segment. Moreover, **child abuse and single-parent household** are salient aspects in the case of minors under 18.

Besides these research findings, it should be noted that some variables are more society-specific or contextual concerning the social situation in the USA and Allegheny, such as the phenomenon of "veterans without stable or supportive housing" (Riden, 2011: 6). As we will see, this group has an important statistical incidence within homeless people living in Allegheny County. Furthermore, several authors have underlined how housing market conditions are increasingly determining the levels of homelessness in the USA (Byrne, et al., 2012; Glynn and Fox, 2017; Nisar et al., 2019). Along these lines, **affordable housing and an increase in foreclosures**, **wages and public assistance** have been underlined (US Interagency Council on Homelessness, 2010). For instance, in Columbia, Riden (2011: 16) has shown how the main causal factors leading to homelessness are: a) Neglect/Abuse; b) Conflict in the Home; c) Delinquent Activities by Youth; d) Parent Substance Abuse/Criminal Behaviors; e) Youth Detention Facility/Without a Place to go; f) Evicted/Could Not Contribute to Rent; and g) Family Lost Housing.

Lastly, other aspects to be considered are those related to how homelessness is treated and examined by experts and public authorities in the country. In particular, in many states, there is a **lack of systematic or quality data about the situation of people living on the streets**. In Columbia: "lack of concrete data able to provide insight on the scope and needs of homeless children and youth, and particularly those who are no longer connected to their families or community of origin, has become increasingly clear." (Riden, 2011: 6).

4- Social setting and homelessness in Allegheny

4.1 Socioeconomic situation in Allegheny

Allegheny County is a county of the USA State of Pennsylvania. The county population is about **1,223,048 people and about 51,7% of them are female** (United States Census Bureau, 2018). In terms of race, around 80,1% of the population are white alone, 13,4% Black or African American alone, 4,1% Asian alone and 2.2% Hispanic or Latino (United States Census Bureau, 2018). It should be noted that only 5.7% of the population was born abroad and that the County has a total of 79,232 veterans.

In 2018, it was estimated that there were **602,414 house units in the County, with 64,9% of occupation rate** (United States Census Bureau, 2018). There is an average of 2.23 people by household (2013-2017) and 86,4% of them have been living in the same house at least during the previous year. Only 7.3% of these inhabitants speak another language than English.

Concerning the health conditions of the Allegheny population, there are **9% of people with a disability under age 65 years** (2013-2017) and **4.9% of people under this age do not have health insurance**. Lastly, in terms of the economic conditions of the population, 64,7% of men and 60.0% of the population over 16 years is in the civilian labor force (United States Census Bureau, 2018). The median household income is $56,333, per capita in the past 12 months $35,280 and the percentage of **people in poverty is 11.2%.**

It is important to note that **poverty and income differences across gender and race groups are very significant**. On the one hand, according to the Census Bureau (2018), in 2017, full-time male employees in Pennsylvania received 1.34 more salary than female employees. Moreover, 2.5% of the population for whom poverty status is determined in Allegheny County, PA (150k out of 1.2M people) live below the poverty line, a number that is lower than the national average of 13.4%. In the same vein, the largest demographic living in poverty is Females 18 - 24, followed by Females 25 - 34 and then Males 18 - 24[9]. On the other hand, according to the Census Bureau (2018), in 2017 the most common racial or ethnic group living below the poverty line in Allegheny County, PA was White (83,991), followed by Black (47,846) and Two Or More (8,596). However, while the black population represented **30,9% of the whole population living in poverty, this race group only represents 13,4% of the whole population of Allegheny County**.

[9] The Census Bureau uses a set of money income thresholds that vary by family size and composition to determine who classifies as impoverished. If a family's total income is less than the family's threshold than that family and every individual in it is considered to be living in poverty.

4.2 Homelessness in Allegheny

To frame the information about homelessness in Allegheny, the methodology applied for counting homeless people should be introduced. The so-called winter **Point-In-Time (PIT) count** is conducted annually in all US cities as mandated by the HUD, which in the case of Allegheny County is complemented with a summer PIT count. As part of this process, data about homeless people is stored in the County's Homeless Management Information System (HMIS), which includes information of each individual facilitated by homeless system providers and by outreach teams that interviewed people in unsheltered locations.

Data is collected following the guidelines yearly produced by HUD, which introduce the federal strategy in the field and include relevant definitions. These definitions determine whether people living in **unstable housing situations are or not defined as homeless** and therefore whether it should be included in the PIT count or not. In fact, "the PIT count does not include those who are in danger of becoming homeless in the near future, living in doubled-up situations, or enrolled in permanent housing programs for the formerly homeless."(Allegheny County, 2019:1).

Following the above methodological criteria, in Allegheny, a total of **783 homeless people**[10] **were identified in 2018**, which represents a decrease of 362 individuals since the previous count in 2018 (Allegheny County, 2019). This number also represents **0,05% of the Allegheny County population only, which is quite below the national average (0,16%)**. Concerning the overall amount of **people living on the street in 2018, 7% (56 people) were unsheltered on the night of the count**, meaning that they were living in a place not meant for human habitation, such as on the street, in a park, in a car or an abandoned building. However, we should also consider that this count does not reflect the total amount of homeless people in Allegheny, since only includes those identified during the night of the census. Furthermore, more individuals considered homeless can be active in the County programs for homeless people. For instance, 7,658 unique people were active in all programs of the County between 1/1/2019 and 9/12/2019[11].

Gender distribution of the homeless population in 2018 shows male predominance. **There were 490 males within the homeless population (63%) and 291 females**, plus 2 people who identified themselves as transgender. Percentage of males prevailed in all the programs implemented by the County and only males were served in Safe Havens. Moreover, **males**

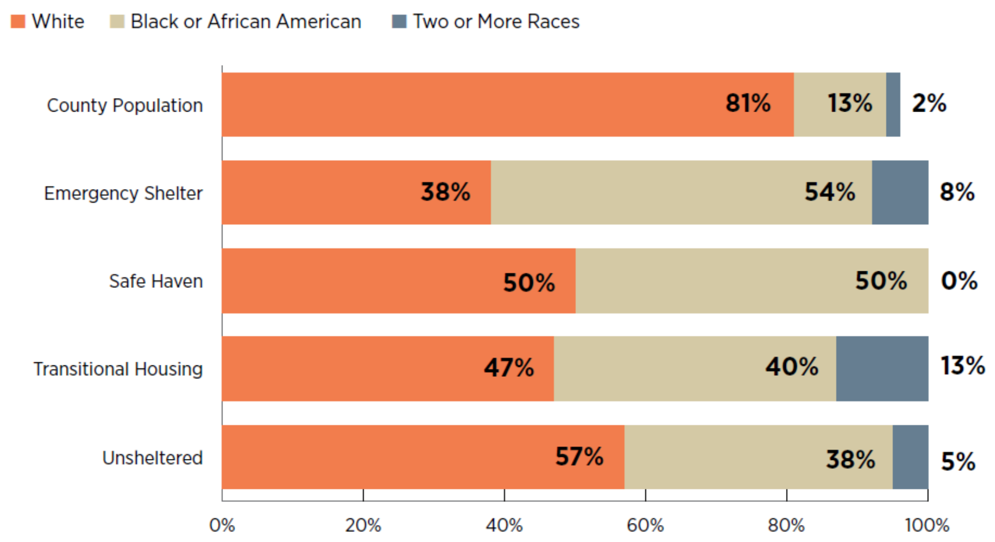
[10] This includes people taking part in the following five programs: 1) Emergency Shelter, 2) Rapid Re-Housing (RRH), 3) Transitional Housing, 4) Safe Haven and 5) Winter Shelter.

[11] Information available at: https://www.alleghenycountyanalytics.us/index.php/2018/07/26/clients-using-allegheny-county-homelessness-programs-interactive-dashboard/.

were 88% of homeless people living unsheltered[12]. This gender distribution shows certain differences with the national balance described above (28,7% female and 70,4% male and only the 0,7% transgender/gender non-conforming).

Regarding **race**, the statistics follow the national trend, with a dominance of Black/African American (395), who are followed by white (320) and mixed-race (52) in 2018. According to the PIT count, **51% of homeless people are Black/African American**, which is highly disproportionate if we take into account that Black/African Americans are only 13% of the Allegheny population according to the Census of 2018 (United States Census Bureau, 2018). The statistic difference can be summarized as follows: "24 out of 1,000 black/African American people in the County were considered to be homeless by the PIT count. This is compared to a rate of less than 1 per 1,000 for white people in the County"(Allegheny County, 2019). The Allegheny report (2019) compares the presence of white, Black/a African American and other races in both the county population and the social services aimed at homeless people, as follows:

Figure 4. Racial demographics of homeless population



Note: Asian, American Indian/Alaskan Native, and Pacific Islander percentages are not included in figure due to small counts.

Source: Allegheny County (2018).

Concerning **age**, Allegheny County has analyzed the local homeless population by grouping them in three age groups, below 18, between 18 and 24 and between 25 and 86. Among the **783 homeless people in 2018, 86%** of the households identified were adult-only households and **14% were households with at least one child under the age of 18**.

[12] It should be noted that there is a certain trend towards reducing female population. In 2017, there were 675 homeless males (59%) and 466 females. Moreover, 3 people identified themselves as transgender, and one person did not identify a gender.

Concerning youth, meaning people "24 years or younger not residing with family members older than age 24", 57 were counted (88% of them were living in emergency shelters), although this group is difficult to detect, since even though not having regular housing, access more frequently to facilities or resources provided by pairs or relatives. It should be noted that **no unaccompanied homeless youth under age 18 were identified.** This is compliant with the DHS's child welfare guidelines, which foster administrations to provide housing for all under-18 unaccompanied homeless children.

As it is possible to see in the Figure 5 below, the same trend was found in 2017 and most of the homeless people are adults without children, although there are a significant proportion of families.

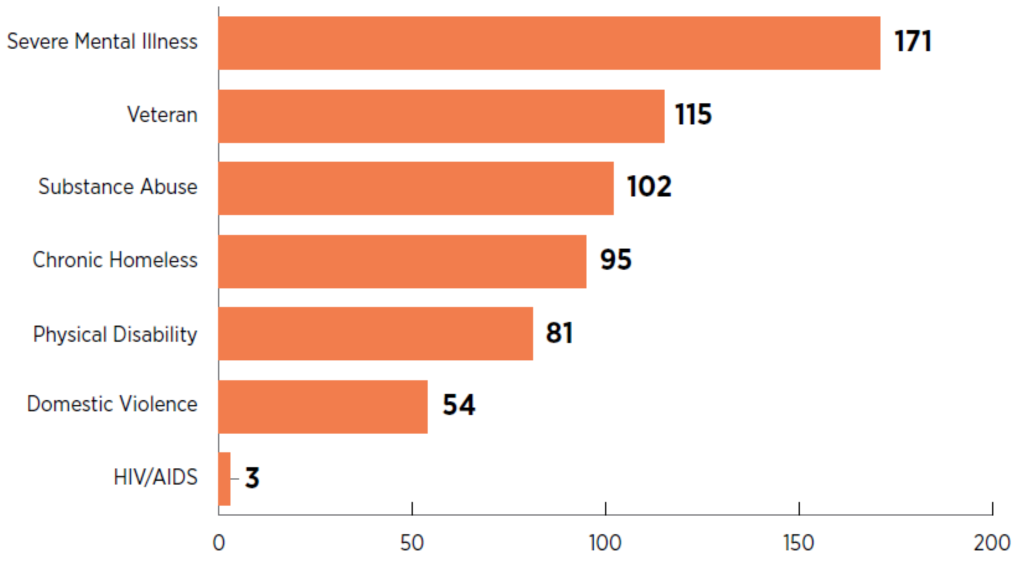
Figure 5. Point in Time Trend by Shelter and Household Type, 2015 through 2015

HOMELESS POPULATIONS	EMERGENCY SHELTER	SAFE HAVEN	TRANSITIONAL HOUSING	UNSHELTERED	TOTAL FOR JAN 2015
JANUARY 2015					
People in families with children	132	0	373	0	505
Individuals without children	308	69	504	38	919
Total	440	69	877	38	1,424
JANUARY 2016					
People in families with children	189	0	212	0	401
Individuals without children	271	54	382	48	755
Total	260	54	594	48	1,156
JANUARY 2017					
People in families with children	216	0	173	0	389
Individuals without children	406	22	275	53	756
Total	622	22	448	53	1,145

Source: Allegheny County (2018).

Regarding the **health condition of homeless people**, Allegheny County has identified that the **largest subpopulation counted are adults with severe mental illness, (171 people),** followed by veterans (115 people), adults with substance use disorder (102 people), people with physical disabilities (81 people), victims of domestic violence (54 people), and people with HIV/AIDS (3).

Figure 6. Homeless Sub populations (Duplicated Counts)



Source: Allegheny County (2018).

As can also be observed in Figure 6 above, a total of 115 homeless veterans were identified in 2018, even though the number of unsheltered veterans went from 7 to 3. Finally, **95 people were identified and classified as chronically homeless population**. As mentioned in the Allegheny County report of 2019, not all of these people can be finally engaged in the permanent housing program.

It should be noted that in 2006, the City of Pittsburgh (Allegheny County Seat), amended an ordinance to crack down on aggressive panhandling after several complaints of downtown business owners. The ordinance (City of Pittsburgh, 2001) is considered one of the harshest on panhandling in the United States and among the actions it punishes is panhandling outside dining areas, admission lines for events, near food dispensing vendors, bus stops and even churches. As soon as it was legislated, the National Commission for the Homeless (2006) considered the ordinance as one of the latest trends in criminalization of homelessness.

5- Algorithmic processing in Allegheny public services and bias

5.1 Automation of public services in Allegheny County and the algorithm for homelessness

Allegheny County has been gradually expanding the automation of some public services management. One of its first systems, the Allegheny Family Screening Tool (AFST), was launched in 2016 (Vaithianathan et al., 2017; Chouldechova, et al., 2018). It uses "statistical modeling to provide hotline screeners with a predictive risk score that shapes the decision whether or not to open child abuse and neglect investigations" (Eubanks, 2016: 14). The system was developed by two scientists, Emily Putnam-Hornstein (University of Southern California) and Rhema Vaithianathan (Auckland University of Technology), with the main purpose of improving the call-screening process in this area (McKenzie, 2018: 559). Basically, social services can now quickly interoperate across several county databases, and enormous amounts of data, to find possible threats for the same children or family. According to McKenzie, (2018:561):

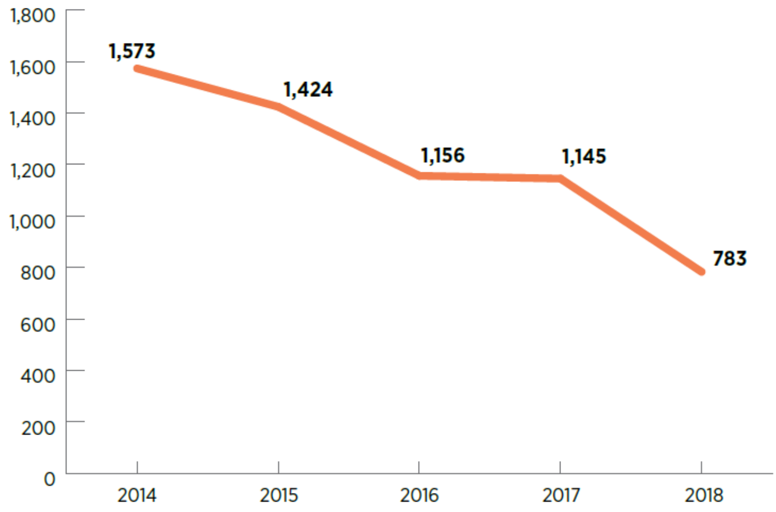
> "the system put in place by Allegheny County has received cautious praise because of the care that has been taken in its implementation, the transparency in its creation, and because the program only calls for investigation, not removal of a child from a family".

According to an assessment conducted by Goldhaber-Fiebert, (2019), this tool may have also **increased the accuracy of the decisions taken to predict risk**. In a similar vein, the system for assigning social services to homeless people seeks to improve the prioritization process. It was also developed by the team integrated by Emily Putnam-Hornstein and Rhema Vaithianathan, following best practices in terms of validation and audit.

In order to analyze the homelessness algorithm implemented in Allegheny, we will first address the conceptualization of the policy goals behind the system. In this context, we will consider the system expected outcomes within the framework of the public resources dedicated to the programs involved in this domain. This programmatic orientation, established by the administration in charge of managing the system, has been considered by the developers of the risk assessment algorithm when setting the ratios for classifying and measuring risk. This contextualization is therefore essential in order to evaluate algorithmic fairness.

The policy of the county has in part been oriented towards **decreasing the number of transitional housing beds and replacing them with permanent housing through *Rapid Re-Housing* and *Permanent Supportive Housing* initiatives** (Allegheny County, 2018). However, the decrease in the number of homeless people in the last few years shown in the graphic below, can be attributed to "(1) a decrease in the number of transitional housing and Safe Haven beds available in 2018, and (2) a lower number of people residing in emergency shelters, particularly Winter Shelters, as compared to 2017" (Allegheny County, 2019: 2)[13].

Figure 7. Total PIT Count, 2014 through 2018



Source: Allegheny County (2018).

Thus, according to Allegheny County (2019), a decrease in the number of homeless people in the County can be mainly explained by the **decrease in the number of people living in Transitional Housing**, which was a strategic policy goal sought by the Allegheny County Homeless Advisory Board. This policy was accompanied by an increase in the facilities for Permanent Housing funded by the HUD and made available by the County. "This was achieved by re-allocating funding from transitional housing programs to Rapid Re-Housing and Permanent Supportive Housing initiatives, which are both considered to be permanent housing and therefore not included in the annual PIT count." (Allegheny County, 2019: 3). Such a process led to a 32% increase in the number of people served in Rapid Re-Housing and Permanent Supportive Housing initiatives **from 2,299 people in 2015 to 3,046** in 2018.
Candidates who **may be subjected to Permanent Housing programs** (and are potentially removed from the homeless category once this transition is completed), should suffer from

[13] Moreover, despite the above mentioned decrease in homeless individuals in 2017, the County has reported an "increase in the number of people served in family shelters" (Allegheny County, 2019: 2).

a **chronic homelessness situation**. The definition of chronic homelessness provided by the Allegheny Council:[14]

> "**[…] requires an individual or head of household to have a disability *and* to have been living in a place not meant for human habitation, in an emergency shelter, or in a safe haven for at least 12 months either continuously or cumulatively over a period of at least 4 occasions in the last 3 years**"

The above definition is similar to the one settled and used by the HUD (2007), which defined chronic homelessness as:

> "either (1) an unaccompanied homeless individual with a disabling condition who has been continuously homeless for a year or more, OR (2) an unaccompanied individual with a disabling condition who has had at least four episodes of homelessness in the past three years."

It should be noted that while the HUD definition involves one of two exclusive options (disability or being unaccompanied) the one used by the Council considers that, to be considered chronic, a homeless person must also be disabled. So, as confirmed by the Allegheny team, this means **a person with no disability cannot be considered chronically homeless**. The Allegheny team has also clarified that this definition was not established by them but is a federal definition of HUD for which they are required to follow and that it is the one used for the algorithmic model.

Furthermore, it should be noted that the policy priorities of the system of homeless continuum of care (CoC) implemented in Allegheny County are expected to be integrated into the new algorithmic system. Currently, these priority groups are:

I. **persons who are chronically homeless,**
II. **families with children,**
III. **transition age youth[15], and**
IV. **veterans.**

The CoC has programs dedicated to serving these populations, and when those programs are full, then these persons can also be enrolled in programs that serve the more general population. According to the County, these populations will continue to receive preference. The basis for this prioritization is the U.S. Department of Housing and Urban Development's

[14] In the Allegheny County report of 2019, the current definition of HUD is determined by the condition of being: "continuously homeless (sleeping in a place not meant for human habitation or living in an emergency shelter) for four episodes in the last three years where the time spent being homeless combined is at least 12 months, or one episode of homelessness that lasts at least 12 consecutive months."

[15] Transition Age Youth are young people, ages 16 to 24, who are at high risk of not successfully transitioning into independent adulthood due to the complexity of their needs, the many challenges they face, and the lack of a support system to assist them.

(HUD) priorities, which has collaborated in the development of the algorithm and funds its associated programs.

To frame the basis of the above prioritization and the possible outcomes of algorithmic processing, the definitions used by Allegheny County for four proxy variables are provided below:

1. **Physical disability**

The Census of the USA defines physical disability as "conditions that substantially limit one or more basic physical activities such as walking, climbing stairs,reaching,lifting,or carrying"[16].

2. **Mental health disability**

The HUD defines severe mental illness as "mental health problems that are expected to be of long-continued and indefinite duration and that substantially impair the person's ability to live independently."

3. **Veterans**

According to HUD, veterans are "any person who served on active duty in the armed forces of the United States. This includes Reserves and National Guard members who were called up to active duty." (Johnson et al., 2017:242).

4. **Chronic substance abuse**

The HUD defines chronic substance abuse as "alcohol abuse, illicit drug abuse, or both that is expected to be of long-continued and indefinite duration and that substantially impairs the person's ability to live independently" (Rockville, 2013).

5.2 Historical bias and hypotheses about algorithmic discrimination

As already explained, the algorithmic system implemented by Allegheny County is aimed at prioritizing possible clients to be benefited by social services while they are in homeless condition or to provide them with Permanent Housing. This risk-ranking system has been trained to measure and classify clients according to an algorithmic statistical model. The methodology used to develop the algorithm is based on its training by using a set of target outcomes and related predictors which provided an overall score for risk. The system should, therefore, be able to identify those individuals who are at a higher risk of homelessness to properly classify them and facilitate their access to a limited amount of housing resources, which should also be adapted to their concrete needs.

Besides these elements determining the efficiency and accuracy of the system, an additional element to be considered is its potential for discrimination based on protected attributes. Along these lines, recipients of CoC Program-funded PSH are required to prioritize otherwise eligible households in a nondiscriminatory manner, following Federal

[16] http://www.census.gov/people/disability/methodology/acs.html.

civil rights laws[17]. Furthermore, other general criteria mandated by the HUD for assessing housing services are aimed to nondiscrimination, such as the ones reflected in Table 4:

Table 4. HUD requirements for assigning access to housing programs and related protected attributes

Conditions	Definition of requirements	Related protected attribute
Few to no programmatic prerequisites to **permanent housing entry**	People are offered permanent housing **does not need to demonstrate sobriety, complete of alcohol or drug treatment, or agreeing to comply with a treatment regimen upon entry into the program**. People are also not required to first enter a transitional housing program in order to enter permanent housing.	Disability
Low barrier admission policies	Permanent supportive housing's admissions policies are designed to "screen-in" rather than screen-out applicants with the **greatest barriers to housing, such as having no or very low income,** poor rental history and past evictions, or criminal histories. Housing programs may have tenant selection policies that prioritize people who have been homeless the longest or who have the highest service needs as evidenced by vulnerability assessments or the high utilization of crisis services.	Property/Disability
Low barrier admission policies	Rapid and streamlined entry into housing – Many people experiencing **chronic homelessness may experience anxiety and uncertainty during a lengthy housing** application and approval process. In order to ameliorate this, Housing First permanent supportive housing models make efforts to help people experiencing homelessness move into permanent housing as quickly as possible, streamlining application and approval processes, and reducing wait times.	Property/Disability

Source: Own elaboration based on Housing First permanent supportive housing.

Taking the above into consideration, the Allegheny algorithm should not only be able to avoid any form of discrimination based on race, gender or disabilities but also integrate proactive mechanisms for facilitating access to housing services in the case of people of **low income, criminal records or living with substance abuse disorders**.

In the following subsections, we summarize the main hypothesis for both problems with the model's predictive accuracy and algorithmic discrimination.

[17] Including, but not limited to the Fair Housing Act, Section 504 of the Rehabilitation Act, Title VI of the Civil Rights Act, and Title II or III of the Americans with Disabilities Act, as applicable.

5.2.1 Hypothesis concerning the accuracy of the model

As we have seen throughout this document, there is a certain alignment between the importance given to **mental health by both the literature and the system as a factor leading to homelessness and the information about the characteristics of homeless people at national and County levels. Nevertheless,** concerning the model, the following questions and issues have been taken into account:

- Does the combined outcome (at least one **inpatient mental health** service in the 12 months following the call, more than four **Emergency Room visits** in the 12 months following the call and at least one Allegheny County j**ail booking** in the 12 months following the call) **effectively measures the risk of homelessness** in the next 12 months? Which group biases could be embedded in this target outcome, e.g., are there sub-groups with a high risk of becoming homeless but that do not tend to use inpatient mental health systems?
    - Is belonging to the **veterans' category correctly prioritized as a predictor**? This element may bias the model taking into account their presence within the whole homeless population in Allegheny (they are the second group after people with mental illness) and considering that they are mainly men.
    - The lack of **direct socioeconomic variables (income, employment)** within the four target outcomes used to modeling the algorithm could undermine the model accuracy, even though mental health or drug abuse can work as proxies of race and poverty.
- Taking into account the very limited availability of services provided, **could people suffering chronic homeless be over-prioritized by the system**? This involves analyzing the ratio defined by the system and its relative alignment with the policy goals behind it.

5.2.1.1. Internal validation of the model

The developing team has already analysed the target outcomes defined for the model were accurate enough. By doing so, they estimated the **marginal effects of receiving Permanent Supportive Housing** on each of the established outcomes. According to this analysis, each of **these target outcomes is lower for those who receive the service**, which suggests that prioritising based on these outcomes would be beneficial.

Other current concerns of the developers include the possibility of using "**substance use services**" as a protective factor rather than as a risk factor. This is because the data source that reports substance use disorder treatments that can be protective, given those individuals are being supported with housing services. According to our analysis below, this outcome could be replaced by a socioeconomic variable, particularly income (through tax or employment information) to remodel the system.

The team also experimented with MH Crisis which is found to be strongly correlated with MH Inpatient and therefore excluded. **Other than that, the outcome "Chronic homelessness" defined as having interactions with shelter or street outreach for four or more months in the following 12 months was not that predictive (AUC was 72%).** Similarly, "Interact with shelter or street outreach" and "Interact with shelter, street outreach, transition, permanent support housing" were not predictive.

Lastly, the team is considering other relevant factors that may be added to the current predictive structure. In this regard, even though they are willing to consider "**experience of violence/sexual violence**" as harm, the lack of administrative data about this variable led to discarding this option. Additionally, the experts pointed out that some previous studies have recognized **gambling, children placing foster care, or children having lower achievements in education as homelessness predictors**. This is in line with Shah et al. (2016), who identified disrupted adoptions and multiple foster care experiences as relevant predictors of homelessness. While gambling is appeared to be collinear with MH Inpatient and Substance use disorder, the team considers worth experimenting with family-specific harms such as **children placing foster care**.

5.2.2 Hypotheses on algorithmic discrimination

On the above basis and after studying the training data and outcomes, we also established some hypotheses concerning group discrimination. Many of these hypotheses have already been tested by the developing team as part of their internal validation of the system.

a) Racial discrimination

As we already discussed, there is an important **historical bias** in terms of the **overrepresentation of Black/African American people** within the Allegheny County homeless population if we consider the overall Black/African American population living in the county. This racial disparity is particularly relevant in the case of male Black/ African American. As noted by the development team, such disparity is explained by other forms of social exclusion that affect this protected group. This includes the already revealed relations between accesses to health treatment; the social capital accumulated by this social group; stigma and worse mental health conditions (Ryan et al., 2006; Fernando and Keating, 2008).
As also reported by the developing team, the DHS "can monitor available data to work toward fair representation". Moreover, fairness in this context does not mean that each race group is represented in all the housing services at the same rate they are represented in the County's population. Still, **this racial disproportionality in the homeless service provision might lead to many forms of discrimination**, for instance, derived from the misrepresentation of Black/African American people due to the correlation of the information on the race attribute with other factors, such as the **amount of information about health for different race groups**.

b) Gender-based discrimination

In terms of gender, there might be a certain i**mbalance between the quality of the data used for training the model for women and men**. This gap could lead to historical bias since it could **worsen the risk assessment for the protected groups**. In particular, **women and LGBTQ people** could be affected due to the lack/quality of information about them and their lower rates within the existing homeless population. This becomes even more important if we consider that, according to Zerger (2008), LGBTQ are often overrepresented within homeless young adults. They are also under a greater risk of homelessness than the heterosexual population.

The **intersections between gender and mental health should be taken into account**. The prevalence of certain mental health conditions in women has been revealed by the literature, which has also found significant geographical variations (World Health Organization (2002). In addition to this, it has been revealed that women are more likely to report and seek treatment for mental health problems (Langan and Pelissier, 2001; SAMHSA, 2002), which has been confirmed for the cases of suicidality and lifetime internalizing disorders (Boyd et al., 2015). Confirming this, conformity to masculine norms has been associated with better mental health (Wong et al., 2017). In this line, it has been pointed out that while women have higher rates of certain mental health disorders, such as major depressive episodes, men present a higher rate of certain substance disorders such as alcoholism (Kessler et al., 1994). Since mental health condition is a key outcome for the algorithmic model, its capacity for capturing these gender differences and nuances should be assessed.

c) Disability-based discrimination

Some health problems such as severe mental or physical illness or limitations are framed as disabilities, leading homeless people to be categorized as chronic. The possibilities of people being categorized as a chronic homeless increase when their disability is considered indefinite and substantially reduce the autonomy of the individual, as shown in the above definitions. This variable is substantially explanatory of homelessness and has an important impact on the model score defined by the algorithm.

- Firstly, this represents a **risk in terms of potential bias against these groups**. Such bias can derive from the data used to identify disabled people. The Allegheny algorithm only uses information and records provided by the Medicare-funded services to assess health interactions concerning mental health and other health variables. The number 0 is used as values for these predictors in the case these data do not exist and self-reported disability is also used. Not having these scores may be problematic since they could hide the impairment and provide less risk (false positives) for these individuals.

- Secondly, the capacity of these data/predictors to **accurately represent the degree and duration of the impairment** should be assessed.
- Lastly, **other diseases**, such as gastrointestinal, neurologic, dermatologic, dental or respiratory diseases should also be considered, given their significant impact among the homeless population (Brown and Steinman, 2013; Feldmann and Middleman, 2003).

d) Property-based discrimination

As indicated above, the importance of factors such as income or poverty levels within the causes leading to homelessness has been pointed out by the literature. Even though the system has been trained on several proxies for social class, socioeconomic factors should be examined to analyze their weight within the model as predictors for risk.

Risks concerning the reproduction of inequality have been found in similar systems. As already revealed by Eubanks (2018), many algorithmic systems implemented by administrations across the US are automating eligibility to social services, in many cases limiting public access to them by those who need them the most. Actually, this has been analyzed for the case of an algorithm used by the city of **Los Angeles to decide who among the homeless people in the city would get housing**. As Eubanks explains, the algorithm gives an advantage to the hardest cases (people struggling with substance abuse or mental illness) and to the easiest ones (people likely to be homeless only for a short time). **Individuals falling between these two categories, lacking job prospects or unmanageable addictions, are a lot less likely to receive any help and end up being homeless for extended periods.**

In the case of the Allegheny algorithm, a set of predictors (including "count of months member was ever in public housing support from Allegheny County Housing Authority (ACHA)") are used to evaluate the risk of homelessness on a property basis. However, these predictors capture the socio-economic condition of the client in the last 12 months only partially. Taking this into account, **the efficiency in predicting the risk of homelessness due to variables such as income and employment should be considered**.

e) Age-based discrimination

The Allegheny County algorithm should be able to **effectively assign social services for housing based on age variables following concrete requirements**. For instance, as we already pointed out, according to the HUD requirements no children below 18 should be living on the street. In this regard, it should be taken into account that children below 18 in families were not measured as children as part of the training data used to train the model. This was because this group is a low risk of homelessness (only 2 cases were part of the training sample, which included 5550 records). Along these lines, in Allegheny, 86% of homeless people are between 18 and 86 age (Allegheny County, 2018).

Still, **risk prediction based on the algorithm should also be able to capture other age specificities. Firstly, age can become a risk factor in terms of vulnerability and a basis for false positives and negatives, for instance in the case of elderly people**.
Secondly, the system should be able to consider as a risk factor the different capitals and housing opportunities held by each age group, both in reality and within the input data. In the following table, we summarize the information about homelessness rates for different age groups and some key references for the main reasons leading to it by each age group.

Table 5. Age groups within the homeless population and prevalent factors leading to homelessness

Age group	% of homeless population (USA) (HUD, 2018)	% of homeless population (Allegheny)[18]	Prevalent factors leading to homelessness (literature)
>18	20%[19]	26% (<17)	Family problems (see adults)
18-24	9%	10%	Trauma and family problems. Domestic violence or sexual violence Mental health (PTSD, anxiety, depressive disorders or self-injuries) Drug and alcohol abuse
24 <	71%	63%	Mental health Substance abuse problems Economic aspects

Source: Own elaboration.

Since **mental health problems seem to increase in number and worsen with age -and time being homeless- and** they importantly determine both the model and the outcomes of algorithmic decisions, the risk for younger segments could be underestimated. According to this, around 36% of homeless people might not be properly represented within the model. The relative prevalence of psychological and drug abuse issues at a young age and economic and mental problems in adults have also implications in terms of the balance between mental health and substance abuse problems.

[18] Based on the clients: https://www.alleghenycountyanalytics.us/index.php/2018/07/26/clients-using-allegheny-county-homelessness-programs-interactive-dashboard/.
[19] Only 5% of all people in unsheltered locations. Only 6,5% were unaccompanied children, at national level (HUD, 2018).

5.2.2.1 Internal validation for groups

The developing team assessed the profiles of individuals who were in the top 10% as per PRM scores in terms of the standard algorithmic fairness criteria for different sensitive attributes including gender, race and disability. The team found race disparities in terms of the percentage of individuals in the top 10% for Jail outcome (19% of more Black individuals). Though, the difference in high-risk groups of combined outcomes is not that significant.

Figure 8. Prevalence top 10% by Race, Gender and Disability

		Race		Gender		Disability (self-reported)	
		Nonblack	Black	Male	Female	Yes	No
	Of Research	40%	51%	49%	51%	84%	16%
Combined PRM Score	Of top 10% as of Combined PRM Score	45%	55%	69%	31%	98%	2%
Jail	Of positive Jail Outcome	41%	57%	66%	34%	88%	12%
	Of top 10% as of Jail PRM Score	40%	59%	76%	24%	91%	9%
	PPV of Jail top 10% for	**53%**	**55%**	**56%**	**48%**	**55%**	**46%**
ER	Of positive ER Outcome	46%	53%	44%	56%	93%	7%
	Of top 10% as of ER PRM Score	49%	51%	46%	54%	98%	2%
	PPV of ER top 10% for	**69%**	**78%**	**72%**	**75%**	**74%**	**70%**
MH Inpatient	Of positive MH Outcome	48%	51%	57%	43%	96%	4%
	Of top 10% as of MH PRM Score	48%	52%	64%	36%	99%	1%
	PPV of MH top 10% for	**64%**	**71%**	**70%**	**63%**	**68%**	**50%**

Source: Center for Social Data Analytics.

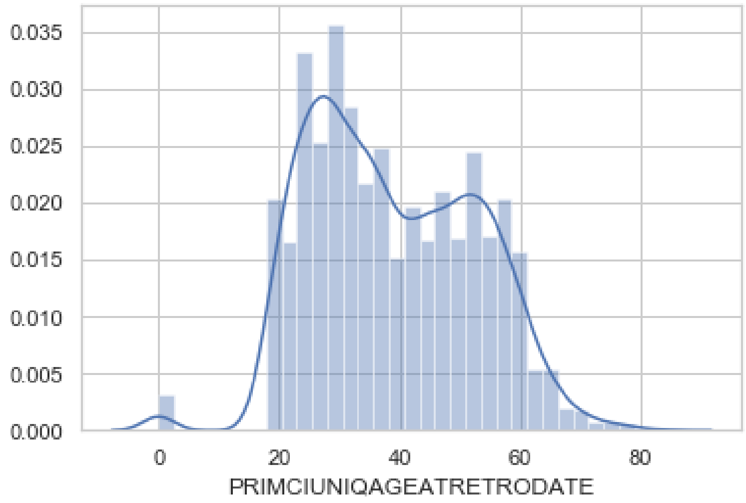
6- Analysis of the dataset composition

In this section, we will analyze the composition of the training dataset. The analysis will focus on some variables such as age, race, gender and some relevant intersections for the analysis of the DI/DT and FNR results. It should be noted that this examination is based on the document titled "2.ResearchData_Homelessness_Allegheny dataset" with size size: 16'793'600 bytes, which corresponds to the audited training data.

6.1 Age groups within the dataset

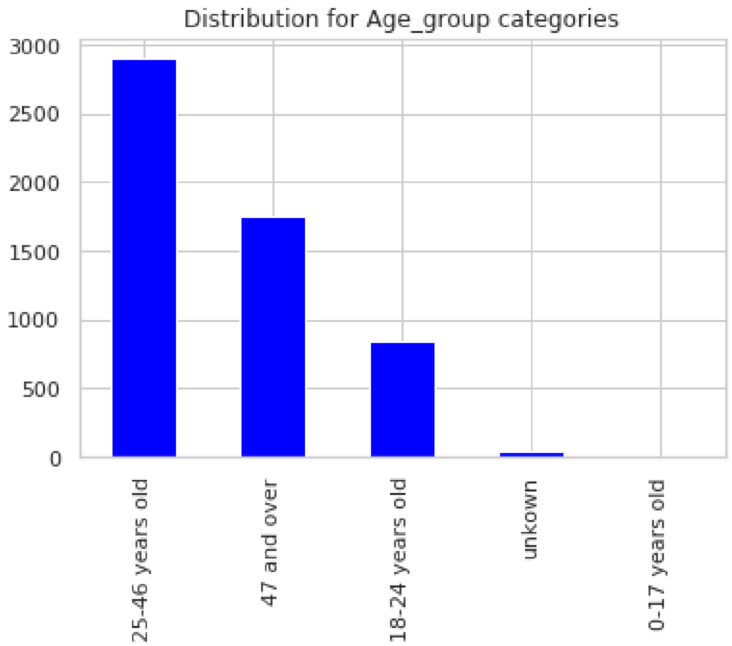
The **median age of the participants at the retro date is 37** (range 0 to 84), (Figure 9). For the gender-related variables, 46 entries have a missing or miscoded gender. **Fifty one percent (2,806/5,504) were female** (Figure 10).

Figure 9. Age distribution at time of retro



Source: Own elaboration.

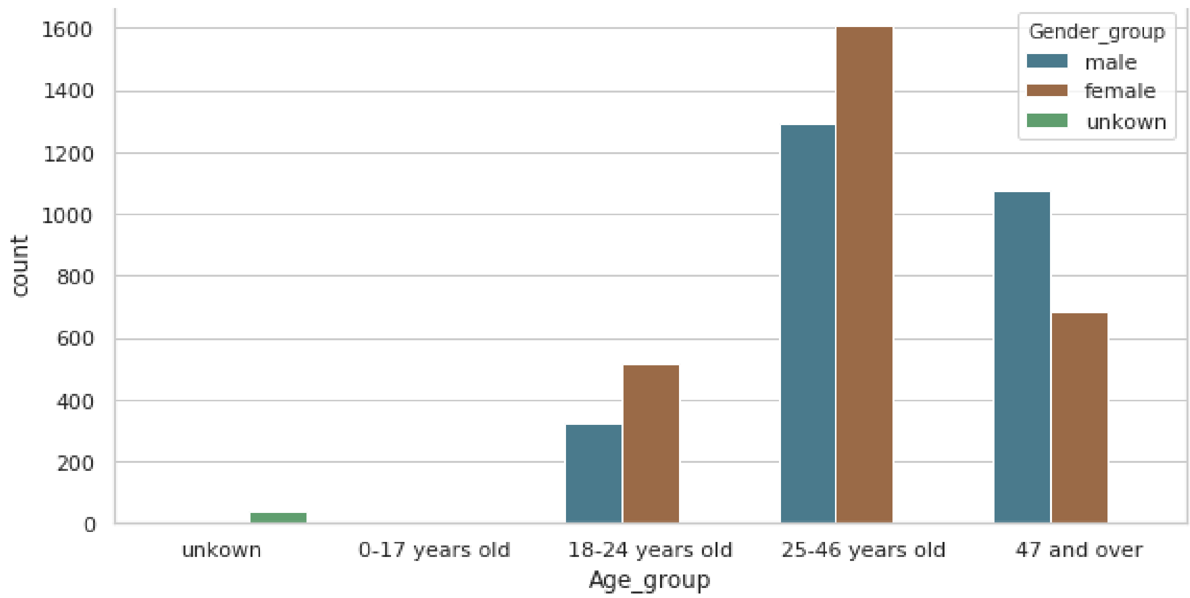
Figure 10. Distribution of age groups at time of retro



Source: Own elaboration.

Among the 5,550 individuals that called Allegheny Link, **the majority (52.3%) belonged to the group of 25-46 years old age group**, followed by 31.7% in the 47 years old and over age category, 15.2% belong in the 18-24 years old category, 0.8% people had unknown age and **there were two individuals in the 0-17 years old category**.

Figure 11. Age distribution by age group and gender at time of retro



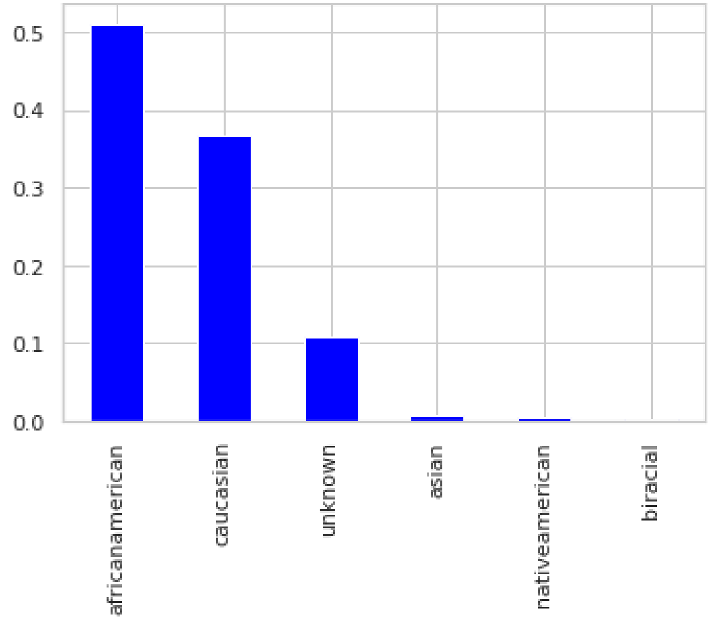
Source: Own elaboration.

As shown in figure 11, in both age groups of 18-24 and 25-46 years old, there are more females that called the Allegheny Link than males; this trend reverses in the 47 years old and over group. This suggests that some intersectional classes that might be considered minorities in the data set (men 25-46, or women 47 and over) could also be considered protected groups.

6.2 Race-based assessment of the dataset

Relative frequencies of racial categories, we observe (Figure 12) several small groups such as native-American or biracial, which suggest that **it might be worth exploring the creation of a model that merges multiple racial categories and has only Caucasian and non-Caucasian as inputs**.

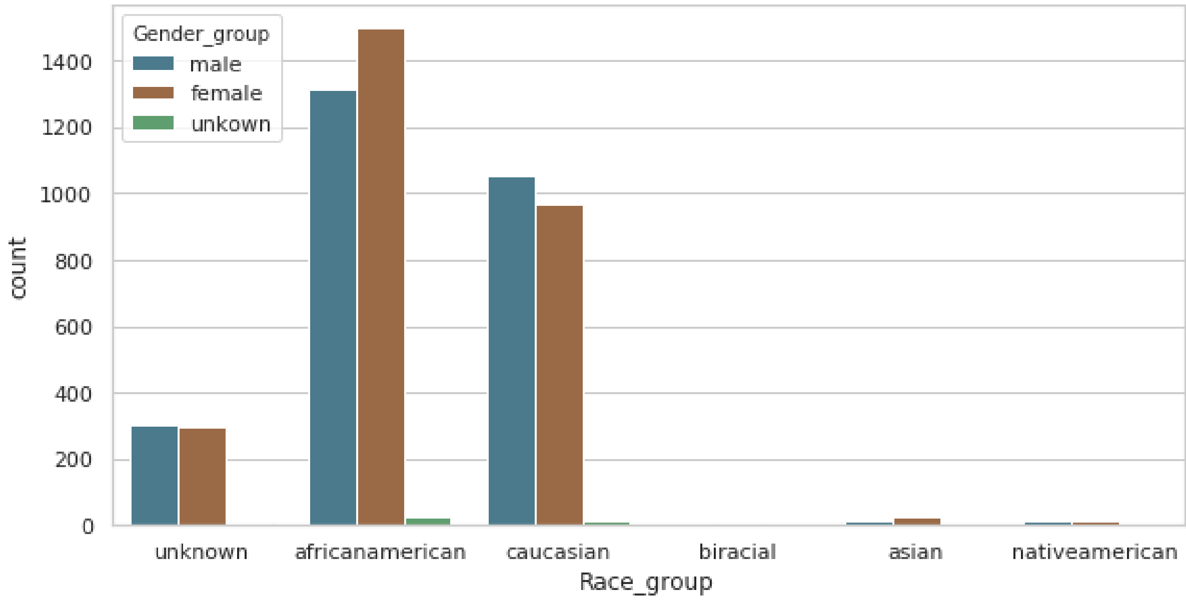
Figure 12. Relative frequency distribution race categories



Source: Own elaboration.

Distribution of race by gender: we do not observe significant disparities in terms of male/female ratio across the two largest racial groups.

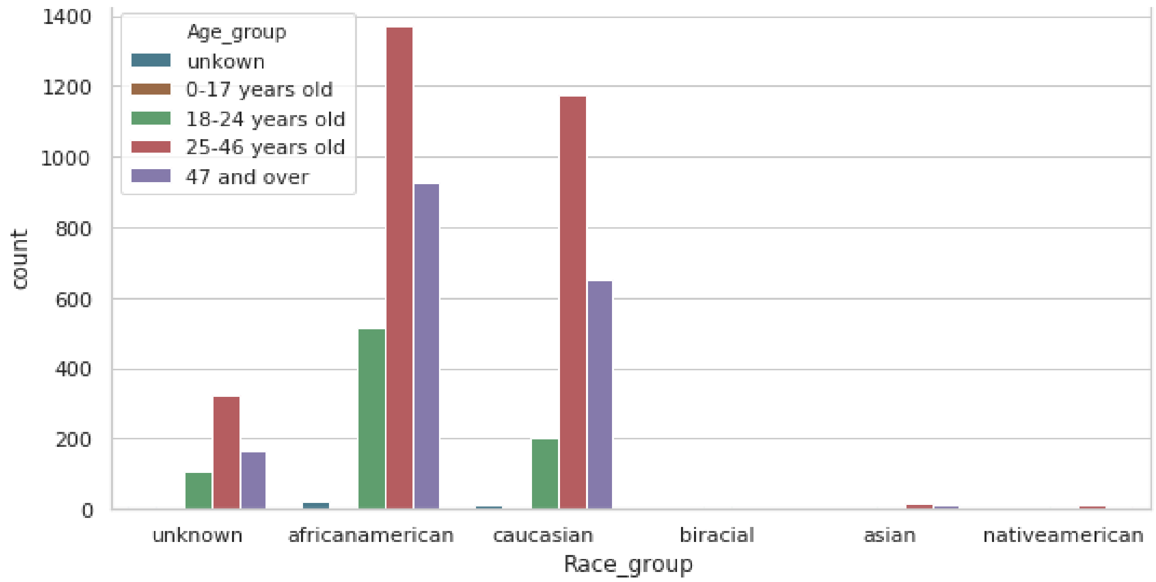
Figure 13. Distribution of race by sex



Source: Own elaboration.

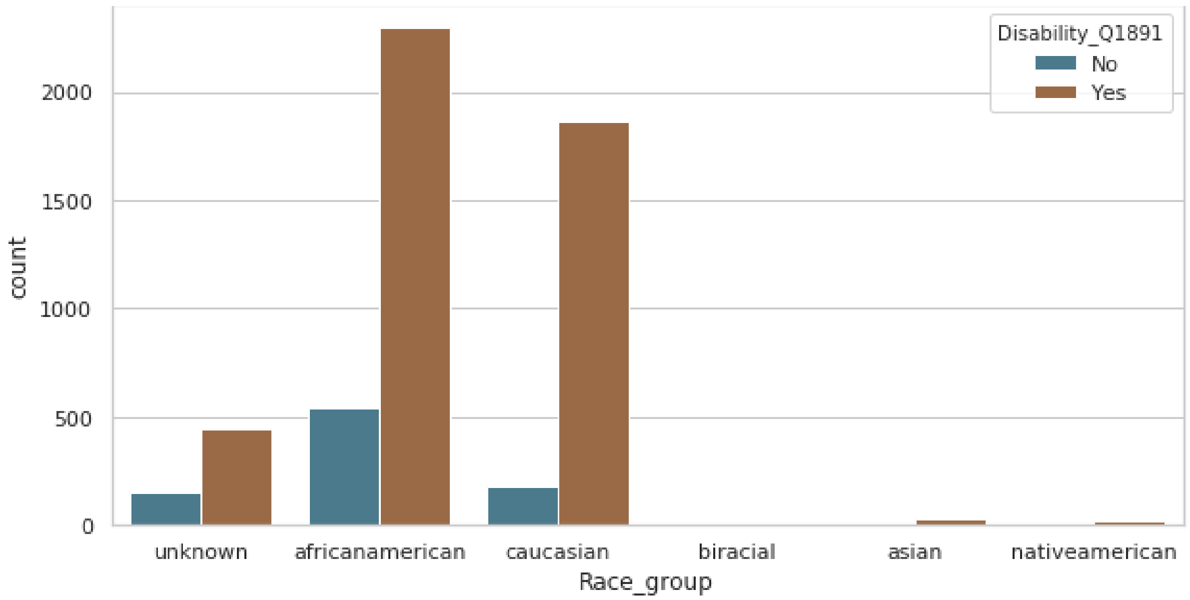
Age groups across race categories: we do not observe significant disparities in terms of age distributions across the two largest racial groups.

Figure 14. Age distributions across racial groups



Source: Own elaboration.

Figure 15. Disability distribution across race categories



Source: Own elaboration.

Overall, there is a large number of individuals with disabilities. **Eighty-four percent of the individuals represented within the data set have disabilities (mental or physical) versus only 15.9% of individuals who do not have any**.

Table 6. Disability by race in selected groups

Disability	African-American	Caucasian
Yes	80.95% (2,294/2,834)	91.37% (1,864/2,040)
No	19.05% (540/2,834)	8.63 % (176/2,040)

Source: Own elaboration.

This same trend is observed across race categories. Notably, the **Caucasian race category presents a higher percentage of individuals with disabilities (91.37%)**. That is ten points more than in the case of African-American. As we note below, the model tends to underestimate risks for people without disabilities.

Table 7. Disability in Black/African-American and Caucasian men

Disability	African-American men	Caucasian men
Yes	81.5% (1,089 /1,337)	90% (963 /1,070)
No	18.5% (248/1,337)	10% (107 /1,070)

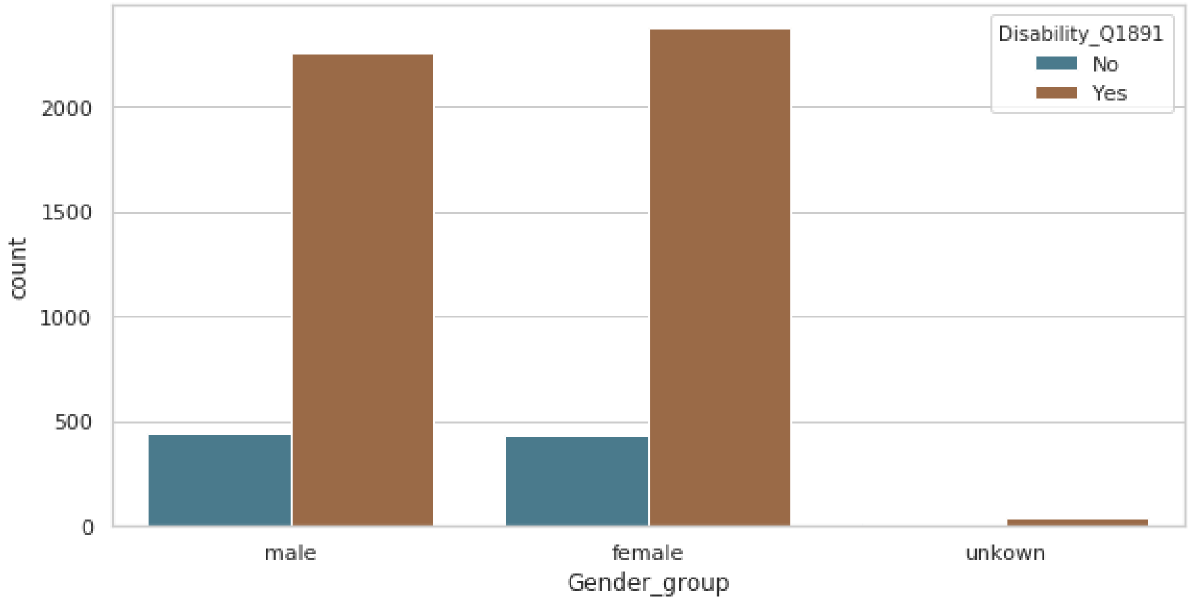
Source: Own elaboration.

In both race categories African-American men and Caucasian men, more than 80% present a disability.

6.3 Gender-based assessment of the dataset

As pointed out above, women are 51% of the Allegheny County homeless population included in the dataset. We can observe almost the same rate of disability among this population than among the males.

Figure 16. Relationship between gender and disability

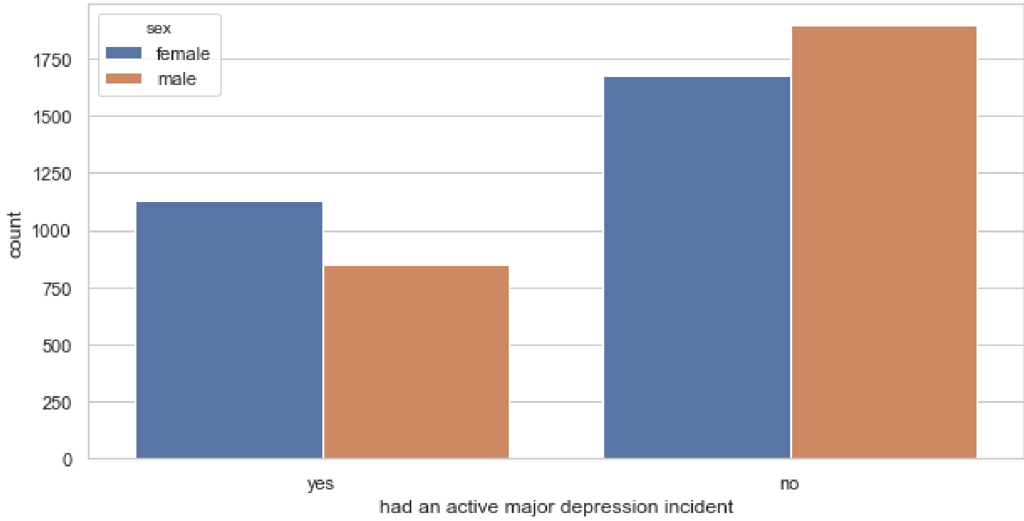


Source: Own elaboration.

Among the individuals who called the Link staff and reported suffering from major depression, 57% were females versus 43% males. It may be possible that inpatient rates are different as WHO has reported[20] that there are gender differences in patterns of help seeking for psychological disorder. For example, women are more likely to seek help from and disclose mental health problems to their primary health care physician while men are more likely to seek specialist mental health care and are the principal users of inpatient care.

[20] Reported at: https://www.who.int/mental_health/prevention/genderwomen/en/

Figure 17. Relationship between reporting an active major depression and gender



Source: Own elaboration.

7. Disparate Impact and Disparate treatment assessment

7. 1. Introduction

The file "**All_results_with_necessaryFlags.csv**" contains the latest version of the dataset that was provided to Eticas (filesize: **1'419'776 bytes**). The features are a mix of categorical, already one-hot-encoded variables (dummy variables), and continuous variables (like *age at retro*). There are different sources integrated into the dataset (data sources: DW, ES/ECAPS, Kids). There are 5550 observations corresponding to unique individuals who called the Link staff, and 964 numerical features for each individual.

Table 8. Before cleaning the dataset:

Family children variable	Value	N (%)
Family with children	1	1,309 (23.6%)
Single	0	4,241 (76.4%)
Total		5,550 (100%)

Source: Own elaboration.

Table 9. Missing data:

Variable	N (%)
Missing gender	43
Missing age	46
Missing race	488

Source: Own elaboration.

Table 10. After removing missing data:

Family_children variable	Value	N (%)
Family with children	1	1,194 (23.8 %)
Single	0	3,829 (76.2%)
Total		5,023 (100%)

Source: Own elaboration.

We divided the dataset into SINGLES (family_children==0) and FAMILY (family_children==1), In order to assess disparate impact and treatment in each dataset separately.

Table 11. Formatted dataset before applying Aequitas method into prediction, outcome and attributes

id	score	Label_value	sex	race	Age_cat	veteran

Source: Own elaboration.

The prediction is based on the "combinedweighted_score_3models_newJailER4MhInp_Decile". In order to apply the Aequitas method, we collapsed the "weighted_score_3models_newJailER4MhInp_Decile" to a binary prediction. We used the thresholds suggested by the CSDA team.

7.2 Analysis of single cases (Family_children==0)

To analyze the Family dataset we established the following cut off:

Table 12. Cut off used in the application of Aequitas methods and interpretation for the data (SINGLE)

Dataset: singles (family_ children==0)	***Combined Weighted Risk Score Deciles : values***	***cut off***	**binarized score for Aequitas**	**interpretation**
Observed outcome MH inpatient Observed ER4 Observed New Jail	1,2,3,4,5,6,7,8	<9	0	low risk
	9,10	>=9	1	high risk

Source: Own elaboration.

We analysed the observed levels of "having at least one mental health related inpatient visit within 12 months" (column named "label_value" corresponding to"'observed_mh_inp_las") versus the prediction from the combined model (column named "score" based on "the binarized weighted_score_3models_newJailER4MhInp_Decile") by protected group.
We established 5 groups of analysis based on their protected attributes (race, gender, age category, disability status, veteran status).

7.2.1 DI/DT assessment by race: outcomes Mhinp, ER4 and New Jail

As it can be observed in Tables 13-16, there is some variability across **race groups**, but the maximum discrepancy between the percentage of people in a group obtaining the "high" combined risk score and the percentage of people in a group for which a negative outcome is observed (MH inpatient, ER4, Jail) is less than ten (10) percentage points. Therefore, in our opinion, the discrepancy is less of a concern.

Concerning the combined weighted score, while high-risk prediction rates are almost the same for the Caucasian and African-American groups, the group "other" has fewer examples and thus exhibits more variability and more differences.

Table 13. Prediction (Combined Weighted Risk Score) by race groups

	Prediction Combined Weighted Risk Score N (%)		
Race groups	**Low**	**High**	**Total N (%)**
Caucasian	1,276 (77.1%)	378 (22.9%)	1,654 (100%)
African-American	1,562 (77%)	467 (23%)	2,029 (100%)
Other	132 (90.4%)	14 (9.6%)	146 (100%)

Source: Own elaboration.

Table 14. Observed MH inpatient risk by race groups

	Observed MH inpatient risk N (%)		
Race groups	**Low**	**High**	**Total N (%)**
Caucasian	1,281 (77.4%)	373 (22.6%)	1,654 (100%)
African-American	1,615 (79.6%)	414 (20.4%)	2,029 (100%)
Other	123 (84.2%)	23 (15.8%)	146 (100%)

Source: Own elaboration.

Table 15. Observed ER4 risk by race groups

	Observed ER4 risk N (%)		
Race groups	**Low**	**High**	**Total N (%)**
Caucasian	1,225 (74.1%)	429 (25.9%)	1,654 (100%)
African-American	1,564 (77.1%)	465 (22.9%)	2,029 (100%)
Other	124 (84.9%)	22 (15.1%)	146 (100%)

Source: Own elaboration.

Table 16. Observed Jail risk by race groups

	Observed Jail risk N (%)		
Race groups	**Low**	**High**	**Total N (%)**
Caucasian	1,344 (81.3%)	310 (18.7%)	1,654 (100%)
African-American	1,588 (78.3%)	441 (21.7%)	2,029 (100%)
Other	121 (82.9%)	25 (17.1%)	146 (100%)

Source: Own elaboration.

7.2.2 DI/DT assessment by gender: outcomes Mhinp, ER4 and New Jail

Similarly than with race, there is some variability between **genders**. The maximum discrepancy between the percentage of people of a gender obtaining a "high" risk score and the percentage of people having that gender that experiences a negative outcome is also less than ten (10) percentage points.

Moreover, discrepancies concerning the probability of getting a high-risk score across genders tend to favour the Male group, except for the ER4 outcome. Even though this risk distribution would favour the most advantaged group, percentage differences in risk assignation are minor.

Table 17. Prediction (Combined Weighted Risk Score) by gender

	Prediction Combined Weighted Risk N (%)		
Gender	**Low**	**High**	**Totals N (%)**
Female	1,275 (80.5%)	309 (19.5%)	1,584 (100%)
Male	1,695 (75.5%)	550 (24.5%)	2,245 (100%)

Source: Own elaboration.

Table 18. Observed MH inpatient risk by gender

	Observed MH inpatient risk N (%)		
Gender	**Low**	**High**	**Totals N (%)**
Female	1,269 (80.1%)	315 (19.9%)	1,584 (100%)
Male	1,750 (78%)	495 (22%)	2,245 (100%)

Source: Own elaboration.

Table 19. Observed ER4 risk by gender.

	Observed ER4 risk N (%)		
Gender	**Low**	**High**	**Totals N (%)**
Female	1137 (71.8%)	447 (28.2%)	1584 (100%)
Male	1776 (79.1%)	469 (20.9%)	2,245 (100%)

Source: Own elaboration.

Table 20. Observed Jail risk by gender.

	Observed Jail risk N (%)		
Gender	**Low**	**High**	**Totals N (%)**
Female	1,359 (85.8%)	225 (14.2%)	1584 (100%)
Male	1,694 (75.5%)	551 (24.5%)	2,245 (100%)

Source: Own elaboration.

7.2.3 DI/DT assessment by age group: outcomes Mhinp, ER4 and New Jail

Concerning differences across **age groups**, the maximum discrepancy between the percentage of people in a group with a "high" combined score and the percentage of people in that group obtaining a negative outcome is also about ten (10) percentage points. As shown in Tables 21-24, the system tends to assign more risk to the group that is more present in the training dataset and within the Allegheny homeless population, 25-46 years. Moreover, the algorithm seems to give more risk to older age groups, which is in line with the dataset and real scenario information provided above, except for the Jail outcome (47 and over have actually a lower chance of going to jail).

Furthermore, the scores from the model for the 0-17 age group cannot be considered valid due to the small number of examples (2). Taking this into account, we consider that this age group requires a specific rule and cannot be handled by model scores.

Table 21. Prediction (Combined Weighted Risk Score) by age group

	Prediction Combined Weighted Risk N (%)		
Age group (years)	**Low**	**High**	**Totals N (%)**
0-17	2 (100%)	0 (0%)	2 (100%)
18-24	425 (85.9%)	70 (14.1%)	495 (100%)
25-46	1,327 (75.5%)	455 (25.5%)	1,782 (100%)
47 and over	1,216(78.4%)	334(21.6%)	1,550 (100%)

Source: Own elaboration.

Table 22. Observed MH inpatient risk by age group

	Observed MH inpatient risk N (%)		
Age group (years)	**Low**	**High**	**Totals (%)**
0-17	2 (100%)	0 (0%)	2 (100%)
18-24	430 (86.9%)	65 (13.1%)	495 (100%)
25-46	1,353 (75.9%)	429 (24.1%)	1,782 (100%)
47 and over	1,234 (79.6%)	316 (20.4%)	1,550 (100%)

Source: Own elaboration.

Table 23. Observed ER4 risk by age group

	Observed ER4 risk N (%)		
Age group (years)	**Low**	**High**	**Totals (%)**
0-17	2 (100%)	0 (0%)	2 (100%)
18-24	401 (81.0%)	94 (19.0%)	411 (100%)
25-46	1,324 (74.3%)	458 (25.7%)	1,782 (100%)
47 and over	1,186 (76.5%)	364 (23.5%)	1,642 (100%)

Source: Own elaboration.

Table 24. Observed Jail risk by age group

	Observed Jail risk N (%)		
Age group (years)	**Low**	**High**	**Totals (%)**
0-17	2 (100%)	0 (0%)	2 (100%)
18-24	396 (80%)	99 (20%)	495 (100%)
25-46	1,352 (75.9%)	430 (24.1%)	1,782 (100%)
47 and over	1,303 (84.1%)	247 (15.9%)	1,550 (100%)

Source: Own elaboration.

7.2.4 DI/DT assessment by disability status: outcomes Mhinp, ER4 and New Jail

As shown in Tables 25-28, in the case of **disability status**, the maximum discrepancy between the percentage of people obtaining the "high" combined risk scores and the percentage of people obtaining a negative outcome is about (10) ten percentage points.

The algorithm shows a tendency to assign higher risk to people with disability over to non-disabled population, which matches the fact that, across all three outcomes, people with a disability indeed experience a negative outcome more often.

Table 25. Predicted (Combined Weighted Risk Score) by disability status

	Prediction Combined Weighted Risk N (%)		
Disability	**Low**	**High**	**Totals N (%)**
No	479 (93.9%)	31 (6.1%)	510 (100%)
Yes	2,491 (75.1%)	859 (24.9%)	3,319 (100%)

Source: Own elaboration.

Table 26. Observed MH risk by disability status

	Observed MH inpatient risk N (%)		
Disability	**Low**	**High**	**Total N (%)**
No	475 (93.1%)	35 (6.9%)	510 (100%)
Yes	2544 (76.6%)	775 (23.4%)	3,319 (100%)

Source: Own elaboration.

Table 27. Observed ER4 risk by disability status

	Observed ER4 risk N (%)		
Disability	**Low**	**High**	**Total N (%)**
No	460 (90.2%)	50 (9.8%)	510 (100%)
Yes	2,453 (73.9%)	866 (26.1%)	3,319 (100%)

Source: Own elaboration.

Table 28. Observed Jail risk by disability status

	Observed Jail risk N (%)		
Disability	**Low**	**High**	**Total N (%)**
No	424 (83.1%)	86 (16.9%)	510 (100%)
Yes	2,629 (79.2%)	690 (20.8%)	3,319 (100%)

Source: Own elaboration.

7.2.5 DI/DT assessment by veteran status: outcomes Mhinp, ER4 and New Jail

In the case of **veterans**, the maximum discrepancy between the percentage of people in a group obtaining a high combined risk score and the percentage of people in that group obtaining a negative outcome is also less than ten (10) percentage points.

Moreover, the system tends to give a high risk score more often to people who are not veteran than to veterans. This agrees with the observation that people who are veteran obtain a negative outcome less often than the rest of the population. Given that veterans are a risk group concerning homelessness, this tendency should be the opposite, but the amount of veteran population in the County could explain this point (second largest homeless population after disabled people).

Table 29. Predicted (Combined Weighted Risk Score) by veteran status

	Prediction Combined Weighted Risk N (%)		
Veteran status	**Low**	**High**	**Totals N (%)**
No	2,710 (76.6%)	830 (23.4%)	3,540 (100%)
Yes	260 (90%)	29 (10%)	289 (100%)

Source: Own elaboration.

Table 30. Observed MH inpatient risk by veteran status

	Observed MH inpatient risk N (%)		
Veteran status	**Low**	**High**	**Total N (%)**
No	2,760 (78%)	780 (22%)	3540 (100%)
Yes	259 (89.6%)	30 (10.4%)	289 (100%)

Source: Own elaboration.

Table 31. Observed ER4 risk by veteran status

	Observed ER4 risk N (%)		
Veteran status	**Low**	**High**	**Total N (%)**
No	2,660 (75.1%)	880 (24, 9%)	3540 (100%)
Yes	253 (87.5%)	36 (12.5%)	289 (100%)

Source: Own elaboration.

Table 32. Observed Jail risk by veteran status

	Observed Jail risk N (%)		
Veteran status	**Low**	**High**	**Total N (%)**
No	2,807 (79.3%)	733 (20.7%)	3540 (100%)
Yes	246 (85.1%)	43 (14.9%)	289 (100%)

Source: Own elaboration.

7.2.6 False negative rate (FNR) by group: outcome MH in-patient

The FNR is the proportion of individuals with a known observed outcome (in this case, had at least one **mental health-related inpatient visit** within 12 months) for which the prediction classifies them as "low risk".We consider there is a *strong* discrepancy when one FNR doubles another, and there is a *matter of concern* if there is a strong discrepancy in which the FNR is higher for an already disadvantaged group.

As it can be seen in the following Table 33, for outcome MH in-patient, there are no strong discrepancies in FNR between men and women (FNR for men is 76% of FNR for women, meaning men are less likely to be underprotected) or between people with disabilities or people without disabilities (FNR for people with disabilities is 63% of FNR for people without disabilities, meaning people with disabilities are less likely to be underprotected.

Concerning age groups, all of them share similar FNR rates. The least protected group is 25-46 years old, which has the most prominent number of samples. Lastly, no reportable differences are identified between races -only in the case of “other races” due to the small number of examples-, nor in veteran- non-veteran groups. In this last case, it should be noted that the rate discrepancy shows a tendency to overprotect the more advantaged non-veteran group.

Table 33. FNR by group (gender and disability)

Group	FNR rate
Men	0.41
Women	**0.54**
With disability	0.45
Without disability	**0.71**
18-24 years old	**0.52**
25-46 years old	0.46
47 years old and over	0.46
Black	0.43
Other	**0.70**
White	**0.49**

Non-veteran (*veteran=0*)	0.46
Veteran (*veteran=1*)	**0.60**

Source: Own elaboration.

7.2.7 False negative rate (FNR) by group: outcome ER4

The FNR is the proportion of individuals with a known observed outcome (four or more **Emergency Room (ER) visits** in the 12 months following the call) for which the prediction is "low risk" . In this case, the system also tends to under-protect the group Women over Men. Concerning disability, the algorithm seems to over-protect the more advantaged group, namely non-disabled population.

In the case of age groups, non-reportable differences are identified. Still it should be noted that the less protected group is 18-24 years old. Similarly that with the inpatient outcome, the algorithm provides a similar FNR for both African-American and Caucasian population but presents a higher rate in the case of other. Lastly, veteran seems again to be under-protected by the system.

Table 34. FNR by group (by gender and disability)

Group	FNR rate
Men	0.41
Women	**0.57**
With disability	0.47
Without disability	**0.80**
18-24 years old	**0.57**
25-46 years old	0.47
47 years old and over	0.49
Black	0.46
Other	**0.86**
White	**0.51**
Non-veteran	0.49
Veteran	**0.58**

Source: Own elaboration.

7.2.8 False negative rate (FNR) by group: outcome New Jail

The FNR is the proportion of individuals with a known observed outcome (At least one Allegheny County **Jail booking** in the 12 months following the call) for which the prediction is classified as "low risk”.

Same as for MH in-patient and ER4, there are no strong discrepancies between groups in general. Rates for Men and Women are almost equal. Still, as observed in the next Table 35, the combined score tends to under-protect the "disability" group less often than the group of people without disabilities.

Table 35. FNR by group (by gender and disability)

Group	FNR rate
Men	0.61
Women	**0.62**
With disability	0.58
Without disability	**0.90**
18-24 years old	**0.72**
25-46 years old	0.59
47 years old and over	0.61
Black	**0.62**
Other	**0.80**
White	0.59
Non-veteran	0.60
Veteran	**0.79**

Source: Own elaboration.

7.3. Analysis of family cases (Family_children==1)

In this section, we present the results for the family cases within the dataset. Metrics were applied to this group following the cut-offs described in Table 36 below.

Table 36. Cut off used in the application of Aequitas methods and interpretation for the data FAMILY

Dataset: FAMILY (family_ children==1)	***Combined Weighted Risk Score Deciles: values***	***cut off***	**binarized score for Aequitas**	**interpretation**
Observed outcome MH inpatient	1,2,3,4,5,6,7	<8	0	low risk
	8,9,10	>=8	1	high risk
Observed ER4 plus (different cut off)	1,2,3,4,5,6,	<7	0	low risk
	7,8,9,10	>=7	1	high risk
Observed New Jail	1,2,3,4,5,6,7	<8	0	low risk
	8,9,10	>=8	1	high risk

Source: Own elaboration.

As with the Single group, we established 5 groups of analysis based on their protected attributes (race, gender, age category, disability status, veteran status).

7.3.1 DI/DT assessment by race: outcomes Mhinp, New Jail and ER4

In this section, we present the results for **Disparate Impact and Treatment by races** corresponding to the outcomes Mhinp (MH in-patient), Jail and ER4. Overall, the algorithm tends to assign slightly more risk to the more advantaged group, Caucasian.

Comparing the combined weighted risk score and observed outcomes, no large (more than ten percentage points) discrepancies between the percentage of people obtaining a "high" combined risk score in a group and the percentage of people in that group obtaining a negative outcome are observed. The larger variation is observed in "other" races, especially with respect to the Jail outcome. These differences may be explained because of the relatively small size of this group, which causes more variability in this statistic.

Table 37. Prediction (Combined Weighted Risk Score) by race groups

	Prediction Combined Weighted Risk Score N (%)		
Race groups	Low	High	Total N (%)
Caucasian	324 (87.1%)	48 (12.9%)	372 (100%)
African-American	706 (90.4%)	75 (9.6%)	781 (100%)
Other	40 (97.6%)	1 (2.4%)	41 (100%)

Source: Own elaboration.

Table 38. Observed MH inpatient risk by race groups

	Observed MH inpatient risk N (%)		
Race groups	Low	High	Total N (%)
Caucasian	342 (91.9%)	30 (8.1%)	372 (100%)
African-American	739 (94.6%)	42 (5.4%)	781 (100%)
Other	40 (97.6%)	1 (2.4%)	41 (100%)

Source: Own elaboration.

Table 39. Observed Jail risk by race groups

	Observed Jail risk N (%)		
Race groups	Low	High	Total N (%)
Caucasian	337 (90.6%)	35 (9.4%)	372 (100%)
African-American	709 (90.8%)	72 (9.2%)	781 (100%)
Other	38 (92.7%)	3 (7.3%)	41 (100%)

Source: Own elaboration.

Table 40. Observed ER4 risk by race groups

	Observed ER4 risk N (%)		
Race groups	Low	High	Total N (%)
Caucasian	303 (81.5%)	69 (18.5%)	372 (100%)
African-American	643 (82.3%)	138 (17.7%)	781 (100%)
Other	37 (90.2%)	4 (9.8%)	41 (100%)

Source: Own elaboration.

7.3.2 DI/DT assessment by gender: outcomes Mhinp, New Jail and ER4

The combined scores, to some extent, seem to be **ignoring differences by gender** that do exist, as they are almost equally likely to give a "high" risk to both men (9.0%) and women (10.7%). However, women actually have a larger risk of becoming MH in-patient and of ER4, and a lower risk of Jail.This means that even tough men and women have a high combined score with a similar probability of 9.0% -10.7%, the probability of MHinp is higher for women (6.8% vs. 2.8%) and there are differences in the other outcomes as well. To some extent, these variations might be explainable by some small sample sizes (e.g., only 6 men had MH in-patient).

Table 41. Prediction (Combined Weighted Risk Score) by gender

	Prediction Combined Weighted Risk N %)		
Gender	**Low**	**High**	**Totals N (%)**
Female	878 (89.3%)	105 (10.7%)	983 (100%)
Male	192 (91%)	19 (9.0%)	211 (100%)

Source: Own elaboration.

Table 42. Observed MH inpatient risk by gender.

	Observed MH inpatient risk N (%)		
Gender	**Low**	**High**	**Totals N (%)**
Female	916 (93.2%)	67 (6.8%)	983 (100%)
Male	205 (97.2%)	6 (2.8%)	211 (100%)

Source: Own elaboration.

Table 43. Observed Jail risk by gender.

	Observed Jail risk N (%)		
Gender	**Low**	**High**	**Totals N (%)**
Female	906 (92.2%)	77 (7.8%)	983 (100%)
Male	178 (84.4%)	33 (15.6%)	211 (100%)

Source: Own elaboration.

Table 44. Observed ER4 risk by gender

	Observed ER4 risk N (%)		
Gender	**Low**	**High**	**Total N (%)**
Female	796 (81%)	187 (19%)	983 (100%)
Male	187 (88.6%)	24 (11.4%)	211 (100%)

Source: Own elaboration.

7.3.3 DI/DT assessment by age group: outcomes Mhinp, New Jail and ER4

In the case of **DI/DT for age groups**, the probabilities of a high combined risk score and of a negative outcome are, to a large extent, aligned. The system follows the same line than with the Single dataset, giving high risk more often to the 25-46 group for all outcomes. The most significant differences in risk low/high-risk assignation happen mostly in small groups, in particular, 0-17 and 47 and over groups. We also observe that, overall, the ER4 probability is higher than what a "high" combined risk score would suggest.

Table 45. Prediction (Combined Weighted Risk Score) by age group

	Prediction Combined Weighted Risk N (%)		
Age group (years)	**Low**	**High**	**Totals N (%)**
0-17	0	0	0
18-24	231 (91.7%)	21 (8.3%)	252 (100%)
25-46	770 (88.9%)	96 (11.1%)	866 (100%)
47 and over	69 (90.8%)	7 (9.2%)	76 (100%)

Source: Own elaboration.

Table 46. Observed MH inpatient risk by age group

	Observed MH inpatient risk N (%)		
Age group (years)	**Low**	**High**	**Totals (%)**
0-17	0	0	0
18-24	239 (94.8%)	13 (5.2%)	252 (100%)
25-46	812 (93.8%)	54 (6.2%)	866 (100%)
47 and over	70 (92.1%)	6 (7.9%)	76 (100%)

Source: Own elaboration.

Table 47. Observed Jail risk by age group

	Observed Jail risk N (%)		
Age group (years)	**Low**	**High**	**Totals (%)**
0-17	0	0	0
18-24	228 (90.5%)	24 (9.5%)	252 (100%)
25-46	783 (90.4%)	83 (9.6%)	866 (100%)
47 and over	73 (96.1%)	3 (3.9%)	76 (100%)

Source: Own elaboration.

Table 48. Observed ER4 risk by age groups

	Observed ER4 risk N (%)		
Age groups (years)	Low	High	Total N (%)
0-17	0	0	0
18-24	206 (81.7%)	46 (18.3%)	252 (100%)
25-46	712 (82.2%)	154 (17.8%)	866 (100%)
47 and over	65 (85.5%)	11 (14.5%)	76 (100%)

Source: Own elaboration.

7.3.4 DI/DT assessment by disability status: outcomes Mhinp, New Jail and ER4

In the case of **disability status**, chances of obtaining a high combined risk score and a negative outcome are not more than 10%. As it is shown in Tables 49-52, the algorithm assigns a higher risk score to people with disabilities, which is aligned with the fact that this group exhibits more actual chances of becoming MH in-patient or ER4. Differences are smaller in the case of the Jail outcome. As with race, it should be taken into account that people with disabilities are preeminent both within the testing sample and the Allegheny homeless population, in comparison with the general population. Here we also observe that, overall, the ER4 probability is higher than what a "high" combined risk score would suggest.

Table 49. Predicted (Combined Weighted Risk Score) by disability status

	Prediction Combined Weighted Risk N (%)		
Disability	Low	High	Totals N (%)
No	221 (96.9%)	7 (3.1%)	228 (100%)
Yes	849 (87.9%)	117 (12.1%)	966 (100%)

Source: Own elaboration.

Table 50. Observed MH risk by disability status

	Observed MH inpatient risk N (%)		
Disability	Low	High	Total N (%)
No	224 (98.2%)	4 (1.8%)	228 (100%)
Yes	897 (92.9%)	69 (7.1%)	966 (100%)

Source: Own elaboration.

Table 51. Observed Jail risk by disability status

	Observed Jail risk N (%)		
Disability	Low	High	Total N (%)

No	212 (93%)	16 (7%)	228 (100%)
Yes	872 (90.3%)	94 (9.7%)	966 (100%)

Source: Own elaboration.

Table 52. Observed ER4 risk by disability status

	Observed ER4 risk N (%)		
Disability	**Low**	**High**	**Total N (%)**
No	195 (85.5%)	33 (14.5%)	228 (100%)
Yes	788 (81.6%)	178 (18.4%)	966 (100%)

Source: Own elaboration.

7.3.5 DI/DT assessment by veteran status: outcomes Mhinp, New Jail and ER4

In the case of **DI/DT for the veterans' group**, chances of obtaining a high combined risk score are larger for veterans. In contrast, their risk of being MH in-patient is equal and of Jail and ER4 is lower for veterans than for non-veterans. We notice, however, that the group of veterans is small and this may explain the variations.

Table 53. Predicted (Combined Weighted Risk Score) by veteran status

	Prediction Combined Weighted Risk N (%)		
Veteran status	**Low**	**High**	**Totals N (%)**
No	1,043 (89.8%)	119 (10.2%)	1,162 (100%)
Yes	27 (84.4%)	5 (15.6%)	32 (100%)

Source: Own elaboration.

Table 54. Observed MH inpatient risk by veteran status

	Observed MH inpatient risk N (%)		
Veteran status	**Low**	**High**	**Total N (%)**
No	1,091 (93.9%)	71 (6.1%)	1,162 (100%)
Yes	30 (93.8%)	2 (6.2%)	32 (100%)

Source: Own elaboration.

Table 55. Observed Jail risk by veteran status

	Observed Jail risk N (%)		
Veteran status	**Low**	**High**	**Total N (%)**
No	1,054 (90.7%)	108 (9.3%)	1,162 (100%)
Yes	30 (93.8%)	2 (6.2%)	32 (100%)

Source: Own elaboration.

Table 56. Observed ER4 risk by veteran status

	Observed ER4 risk N (%)		
Veteran	**Low**	**High**	**Total N (%)**
No	954 (82.1%)	208 (17.9%)	1,162 (100%)
Yes	29 (90.6%)	3 (9.4%)	32 (100%)

Source: Own elaboration.

7.3.6 False negative rate (FNR) by group: outcome Mhinp

The FNR is the proportion of individuals with a known, negative observed outcome (had at least one mental health related inpatient visit within 12 months) for which the prediction classified the individual as "low risk".

As shown in Table 57 below, in general, there are no strong differences in the FNR by a protected group, except for the case of veterans in which the FNR for non-veterans is about half of the FNR for veterans.

This means that veterans are more likely to be under-protected than non-veterans. Ideally, this should be neutral, or veterans should be less likely under-protected by the system. In the same line, also younger people are more likely than people of 47 years or older to be under-protected by the system. The less protected group is 18-24 years old. Moreover, the algorithm tends to under-protect women over men. Concerning race, the algorithm follows the same line than with other outcomes and provides almost the same protection to Caucasian and African American and more to the group others.

Table 57. FNR (gender, age group, race and disability, veteran)

Group	FNR rate
Men	0.33
Women	**0.57**
With disability	0.54
Without disability	**0.75**
18-24 years old	**0.62**

25-46 years old	0.56
47 years old and over	0.33
Black	0.52
Other	**1.00**
White	**0.57**
Non-veteran	**0.56**
Veteran	0.00

Source: Own elaboration.

7.3.7 False negative rate (FNR) by group: outcome New jail

The FNR is the proportion of individuals with an observed negative outcome (at least one Allegheny County **Jail booking** in the 12 months following the call) for which the prediction classified the individual as "low risk".

Regarding the FNR by a protected group concerning the outcome Jail, the algorithm tends to protect more often people with a disability in relation to people without disabilities. In this case, the discrepancy is substantial, but it is not worse for the most vulnerable group (those with disabilities). These results are, therefore, not a matter of concern since this score should be neutral or higher risk should be assigned for people with disabilities.

In the other groups, differences are smaller except for “other" races which might be due to the small size of this group. Men are slightly more under-protected than women and results concerning race do not show significant discrepancies. Lastly, veterans are highly more protected than non-veterans which do not go against the fairness framework defined above.

Table 58. FNR (gender, age group, race and disability, veteran)

Group	FNR rate
Men	**0.70**
Women	0.61
With disability	0.59
Without disability	**0.94**
18-24 years old	**0.75**
25-46 years old	0.60
47 years old and over	0.67
Black	**0.68**
Other	**1.00**

White	0.51
Non-veteran	0.65
Veteran	**0.00**

Source: Own elaboration.

7.3.8 False negative rate (FNR) by group: outcome ER4

The FNR is the proportion of individuals with an observed negative outcome (Four or more **Emergency Room (ER) visits** in the 12 months following the call) for which the prediction is "low risk".

Differences in FNR are not large in general, except for the group of veterans, who are more under-protected at about half the rate of non-veterans. People with disabilities are also less likely under-protected than people without disabilities. Women are more under-protected than men, and as with the single group, the group 18-24 is more under-protected than others.

Table 59. FNR (gender, age group, race and disability, veteran)

Group	FNR rate
Men	**0.54**
Women	0.62
With disability	0.58
Without disability	**0.76**
18-24 years old	0.67
25-46 years old	0.60
47 years old and over	0.45
Black	**0.64**
Other	**0.75**
White	0.55
Non-veteran	0.62
Veteran	**0.33**

Source: Own elaboration.

8- Analysis

This document firstly addressed the algorithmic and policy goals of the service developed in Allegheny aimed at predicting the risk of homelessness. In this framework, two elements must be considered from an algorithmic fairness perspective. First, the system should be able to correctly classify individuals at high risk of homelessness, thus not reducing their social protection or assigning them to an inappropriate social program. On the other hand, this classification should be carried out without giving a low risk of homelessness to individuals who need assistance and belong to protected groups, such as people with disabilities or otherwise disadvantaged people.

The above requirements concerning accuracy and non-discrimination are based on both applicable legal texts (including the Fair Housing Act, the Rehabilitation Act, the Civil Rights Act and the Americans with Disabilities Act) and the policy of the HUD. In terms of policy goals, the system should be capable of allocating public housing resources effectively -from both human resources and an economic standpoint-, legally -in line with the applicable legal requirements on human rights, as well as the standards established by the HUD- and following a pre-established policy strategy. In this regard, Allegheny County explicitly aims at decreasing the homeless population while improving the alignment of housing supply and demand. These policy goals require the improvement of the existing prioritization methods in terms of speed, which has to be accomplished in a manner that keeps accuracy levels as high as possible.

In this section, we will summarize the results of the algorithmic impact assessment conducted on the system. We will also contrast the system's theoretical basis and the outcomes of the processing results against the above technical and policy objectives. To do so, we will cover four different levels of analysis. Firstly, we will briefly re-describe the dataset composition and analyze the overall accuracy of the system, considering data quality and its potential impact on the modeling process. Secondly, we will compare the historical bias with the findings of the assessment in order to spot possible inaccuracies, bias or actual discrimination. We will also analyze these results by carrying out a risk assessment concerning our preliminary hypotheses. Thirdly, we will provide a desirability analysis based on these findings. Lastly, we will provide recommendations for future redesign and implementations.

8.1 Dataset composition, model training and overall accuracy in risk assignation

8.1.1 Protected groups within the dataset

As we already pointed out, algorithmic bias can derive from unrepresentative or incomplete training data or the reliance on flawed information that reflects historical biases. The data used to fine-tune the algorithmic model may be biased, which negatively affects the machine learning process, since ML systems learn to identify factors in the data that enable them to make predictions concerning an expected outcome. To address this issue, in this section we will describe the distribution of the homeless population present in the training data across four fundamental protected categories, age, disability, gender, and race.

The training sample was composed of 5,550 records of individuals who called Allegheny Link. The majority of them belong to the 26-45 age group (53.2%) and the **vast majority of them are above 26 years old (86,9%)**. This is in line with the socio-demographic configuration of homeless people in Allegheny according to its PIT count.

The **majority of the clients reflected in the sample are mentally or physically disabled (84%)** versus only 15.9% who are not. This is consistent with the existing situation in Allegheny and at the national level, where disability rates among homeless people are above 80%.
In terms of gender, the structure of the training sample is **quite proportionate (49% male and 51% female)**. However, this does not completely reflect the gender distribution of homeless people in Allegheny, which has a majority of men (63%). The underrepresentation of men (49% in the sample vs 63% in the homeless population) theoretically has the potential of harming the accuracy of the system.

Lastly, concerning race within the records used to train the model, while there is a very low presence of native-American or biracial, the data shows that **54% Black and African American and 39% Caucasian people called the Allegheny Link**. This is only 3% below the number obtained in 2018 during the PIT count in Allegheny for Black/African Americans (51%) (Allegheny County, 2018).

8.1.2 Intersectional structure of the training data

As suggested by Foulds et al., (2018), to properly address bias and discrimination within algorithmic processing **an intersectional definition of fairness must be applied**. While belonging to a protected group could represent a certain advantage or impairment in regards to other social groups, belonging to more than one protected group can also make a difference in terms of both the efficiency and the potentially discriminatory character of the system.

Taking this into account, we have examined the Allegheny dataset and identified some intersections between protected attributes that are relevant for framing discrimination within the studied case. In particular, given the model design and according to our hypotheses, two intersections are relevant for its evaluation: **Disability and Gender, and Disability and Race**.

Firstly, **disabilities are more present among Caucasian (91.37%)** than Black/African American people (80.95%). Rates are quite similar in the case of Caucasian women (90%) and Black/African American women (81.5%), which means that there are no remarkable differences between genders. Moreover, **most of Black/African American people integrated into the sample are women.**

Secondly, we found that **the algorithm tends to under-protect more often women than men,** although not by a large margin.

8.1.3 Overall identified accuracy in risk assignation

In general, the algorithm **seems to accurately assess those social groups** who, according to the statistics about homelessness reflected in the training dataset, present the highest risk of homelessness, including people with disability. This general trend is manifested in the results of **both DI/DT and FNR** assessments. Some variations concerning the different outcomes used by the model are identified, but they all are in line with this trend. The provision of higher risk to men than women also seems to be in line with the actual real risk factors based on the literature and also statistics about homelessness in the County. Lastly, the higher risk assigned by the algorithm to the 25-46 age group corresponds to the statistics about the age composition of the Allegheny homeless population.

Indirect evidence of **lower accuracy** can be found in the **assignment of risk in the case of veterans** for the Single dataset, where this group presents less risk than non-veterans. Instead, veterans in the Family datasets show the opposite trend, being under more risk than its dichotomic group; we should take into account that the number of veterans with children in the family is quite low (with just 32 observations) and making inferences may be hard. On the one hand, we should consider that, regarding the health condition of homeless people, veterans are the second largest group in the County. On the other hand, this contradiction seems not to be in line with the actual risk that these groups might have under these two different conditions.
Along these lines, even though the provision of risk by race seems to be equally distributed across groups, the FNR rate **tends to be slightly larger for Black/African American,** with the

consequence of under-protection, in the case of the Jail outcome. This last result could be explained by the more significant presence of Black/African American population in prison, although the difference between African-American and Caucasian is only 4% in Allegheny County[21]. The results, therefore, go slightly against the statistical evidence of race as a risk factor when it comes to the slight prominence of Black/African American within the Allegheny County homeless population. Even though taking into account this information about veterans and race, it should be noted that differences in risk attribution showing potential inaccuracy in the algorithm performance are minor in all cases, except for the race "other", which may be explained by the less amount of samples for this group within the training dataset.

8.2 Bias and potential discrimination within the system

Social biases and other socio-technical issues are often integrated within the training data used to train the models even if preventative measures are taken when sampling and conducting feature selection (Suresh and Guttag, 2019). Along these lines, already existing bias described in the previous section may harm the model and favor algorithmic discrimination. This issue has been evaluated by the development team, which found biased outcomes against Black/African American for some outcomes. Moreover, other common forms of bias, such as measurement bias, are relevant for the Allegheny algorithm. In this case, biased results are explained by how data was collected, detected or measured (for instance, through Medicare). In this section, we will summarize our findings concerning algorithmic discrimination and frame them within the context of these and other sources of bias.

Our analysis of algorithmic bias, using the definitions in sections 2.1 and 2.2, has been focused on group discrimination, and examined carefully whether the criterion of independence is satisfied in cases where we expect outcomes to be independent, by **looking at differences in the probability of obtaining a "high" combined risk score in cases where the probability of obtaining a negative outcome** (MHinp, ER4, Jail) are similar, and to some extent sufficiency when these probabilities are different, as we would expect that the combined risk score is higher with higher probability for groups that experience a negative outcome at a large rate. **Differences are calculated as a percentage increase; when above 50% they are considered problematic, at least from the theoretical standpoint**. We have also looked at the criterion of separation by looking at false positives and false negatives across groups.

[21] Information available at: https://www.alleghenycountyanalytics.us/wp-content/uploads/2016/06/Changing-Trends-An-Analysis-of-theACJ-Population-FINAL.pdf.

In the following Table 60, we can observe relevant information about the gender, race, disability and age composition of the homeless population in Allegheny, as well as findings for each of these groups within the dataset used to train the algorithm. This data is compared with the main results found concerning disparate impact/treatment and false positive/negative analysis for each of these attributes. Differences, or percentage increases, in the FNR between groups are obtained as (1 – (high FRN / low FRN)).

Table 60. Homelessness in Allegheny by group, distribution within the dataset and bias

Homelessness in Allegheny	Dataset composition	DI/DT analysis[22]	FNR analysis
GENDER: Most of the homeless people in Allegheny are **men (63%)[23]**	**49%** of homeless people are men.	➔ Single: men are assigned a high risk score more often than women, and tend to experience a larger MH in-patient and Jail risk than women, but a lower ER4 risk. ➔ Family: men and women are assigned a high risk score at about the same rate, but have a lower MH in-patient risk and a higher Jail risk.The ER4 risk is the highest. Larger differences are smaller than 10 percentage points.	➔ Single: Women are more likely under-protected than men for the MH in-patient and ER4 risk. Larger differences are about 30%. ➔ Family: Women and men are under-protected at about the same rate. Larger differences are about 15%.
RACE: Most of the homeless people in Allegheny are **Black/African American (51%).**	**54%** of homeless people are Black/African American.	➔ Single and Family: No relevant differences- across outcomes. Large variability is observed in the "Other" races and might be because this is a smaller group. The larger differences are smaller than 10 percentage points.	➔ Single: for MHinp and ER4 Caucasians are more likely underprotected than African American. In the case of New Jail, African American are more likely to be underprotected than Caucasians. Larger differences between these two are about 14%. The greater differences, though, are with "other" races. In this case, larger differences are about **87%**, and might be because this is a smaller group. ➔ Family: Caucasians tend to be less likely under-protected than African American. There are also some differences with "other" races. Larger differences are about 15%-25%.

[22] In the case of Families, the ER4 outcome has a different cut off than that of MHinp and Jail.
[23] 88% of them are living unsheltered.

DISABILITY: The largest subpopulation within the homeless people in Allegheny are **disabled people** (including severe mental illness, chronic homeless and physical disability**- 41%**[24]).	**84%** of people represented individuals with disabilities.	➔ Single and Family: A high risk score tends to be given more often to people with disabilities, which is in line with the fact that the group of people with disabilities experiences higher risk of a negative outcome (MHinp, ER4, Jail). For Singles, larger differences are about 17-18%. For Families, larger differences are smaller than 10%.	➔ Single: people with disabilities are under-protected less likely than people without disabilities. Larger differences are about 35%. ➔ Family: people with disabilities are less likely to be underprotected than people without disabilities. Larger differences are about 40%.
AGE: Most of the homeless people are **adults (86% are <25)**	**86,9%** people are < 26 The majority (53.2%) belongs to the group between 26-45 years old.	➔ Single and Family: A high risk is assigned more often to the 25-46 age group. ➔ The 0-17 age groups may not be considered valid so as to shape the model, given the small sample it represents within the training dataset. For Singles, larger differences are about 10 percentage points. For Families are smaller than 10 percentage points.	➔ Single: for MHinp the group of 25-46 is more likely to be underprotected than the other groups. For ER4 and New Jail, the group of 18-24 is more likely to be underprotected than older people. Larger differences are about 21%. ➔ Family: for MHinp and ER4, the group of 47 and over is less likely to be underprotected than younger people. Larger differences are about 40%.
VETERAN: Regarding the health condition of homeless people, veterans make the second largest subpopulation (**15% are veterans**)		➔ Single: The system tends to give a high risk score more often to people who are not veteran than to people who are veteran. This agrees with the observation that people who are veteran obtain a negative outcome less often than the rest of the population. ➔ Family: Chances of obtaining a high combined risk score are larger for veterans. In contrast, their risk of being MH in-patient is equal.In the case of Jail and ER4 the risk is lower for veterans than for non-veterans. For Singles, larger differences are about 12%. For families, are smaller than 10%.	➔ Single: veterans are more likely to be underprotected than non-veterans. Larger differences are about 30% ➔ Family: veterans are underprotected at about half the rate of non-veterans. Except for the case of New Jail, where FNR is 0 for veterans.

Source: Own elaboration.

[24] It should be noted that this 41% is an estimation based on the Allegheny PIT count of 2018, which does not reflect the overall percentage of disabled people.

Based on the information reflected in Table 60 above, some elements concerning the **risk assignation** of the algorithm must be stressed:

➔ **DI/DT analysis**

There are no significant discrepancies between combined weighted and observed risk by outcome across **genders** (no more than 10%). The same happens across **races** (no more than 10%). The distribution of these groups within the dataset, which have a slight prominence of Black African American and Men, may explain that these two groups are assigned with more risk and are also more protected by the algorithm. No relevant differences have been found for the Families **disability** status. There is a **17-18%** difference for the Single disability status. In this case, the discrepancy is large but is worse for the least vulnerable group (those without disabilities).

Concerning the variable **age**, there are no reportable discrepancies. However, the low level of representation of some age groups in the dataset, particularly below 17 and above 45, should be taken into account when evaluating and monitoring the accuracy of the system. In the case of the Single groups of **veterans** and non-veterans, the system tends to give a high-risk score more often to people who are not veterans. Differences are about 12%. For the veteran/non-veteran Families groups, there are no reportable differences.

Moreover, the following results and potential forms of bias are identified based on the **FP/FN rates**:

➔ **FP/FN for single dataset**

Overall, people without disabilities are more likely to be under-protected by the system (35% more often as the larger difference) than people with disabilities. As already mentioned, this is expected due to the importance of disability as a risk factor for homelessness.

The algorithm tends to **under-protect women** more often than men for the MH in-patient and the ER4 risk with differences bigger than 30% (FNR of 0.54 women vs 0.41 men for MH in-patient which makes a difference of 32% in FNR, FNR of 0.57 women vs 0.41 men for ER4 which makes a difference of 39%). Such discrepancy is not justified either by the composition of the homeless population in Allegheny or regarding the training dataset.

There are some **racial disparities among groups**. Caucasians are more likely be under-protected than African American (FNR of 0.43 African American vs 0.49 Caucasian for MH in-patient which makes a difference of 14%, FNR of 0.46 African American vs 0.51 Caucasian for ER4 which makes a difference of 11%), except for the case of New Jail, where African Americans are more likely to be under-protected (FNR of 0.62 African American vs 0.52

Caucasian for New Jail which makes a difference of 19%). This fact is neither explained by the jail population composition, which shows almost the same presence for both Caucasian and Black African American nor by the overall importance of race as a risk factor[25].

Results for Mhinp show that the **age group from 25-46 is more likely to be under-protected** than other age groups, which is in line with its essential weight within the training dataset and the homeless population. However, for ER4 and New Jail, the group from 18 to 24 is slightly more likely to be under-protected than older groups (FNR 0.58 vs 0.46-0.60 of other age groups for ER4, FNR 0.69 vs 0.60-0.61, which are differences of about 15%).

Finally, the Single analysis shows that **veterans are more likely to to be under-protected** than non-veterans for MH in-patient and Jail with differences above 30% (FNR 0.60 vet/0.46 non-vet for MH in-patient with a difference of 30%, FNR 0.79 vet/0.60 nonvet for Jail with a difference of 32%).

➔ **FP/FN for family dataset**

Concerning FNR for **Families**, in the case of **disabilities**, there is a difference of about 30% among the group with disabilities and the group without disabilities, with the latter more often under-protected (FNR of 0.75, 0.76, 0.94 for people without disabilities in outcomes MH in-patient, ER4, Jail, FNR of 0.54, 0.58, 0.59 for people with disabilities in the same outcomes, which yields a difference of 39%, 31% and 59%, respectively).

Both **genders**, women and men, are under-protected at about the same rate, with non-significant differences.

African American are more likely to be under-protected than Caucasian and "other" **races** also present differences with these two. Larger differences are about 15%-25%. Since there are very few people in the "race = other" group, it is more probable to see variability in this group in terms of false positives, for example. Therefore, the observation that there is a difference of 87% can be explained by the small number of people in this group.

There is also a difference of about 40 percentage points in favor of the **age** group of 47 and over (MHinp and ER4), which is less likely to be under-protected than younger people.

It should be noted that the **only group** of the study (both for Singles and Families) that present an **FNR difference ≥ 50%** is that of **veterans** (Families). In this case, non-veterans are more likely (50% difference) to be under-protected than veterans, they also have a New

[25] It should be taken into consideration that while the amount of Caucasian and Black/African American prisoners is almost the same around 80,1% of the population
are white alone in the County.

Jail and MH Inpatient FNR of 0 for veterans. This may be explained because none (or a very small percentage) of the (32) veterans in the (family) training data have ever been in jail (2 observations from the dataset), had a MH Inpatient (2 observations from the dataset) or ER4 (3 observations from the dataset).

In general, these results show that there are no reportable algorithmic discrimination results except for the **disability** status, which we would not consider discriminatory, given that it favours the most vulnerable group (those with disabilities).

As for the FNR in the case of singles, there is a **potential bias of the system against the group of veterans**, something that should be monitored. Instead, the FNR for Family shows that veterans are overprotected, which is in line with our approach to fairness.

A question to be addressed is that while **women are slightly (less than 10 percentage points) more likely to be under-protected than men with respect to ER4 and MH in-patient risks**, they are under more risk in these categories when they are in Family, which do not seem to have a clear explanation.

Lastly the low representation of some groups in the training data may be affecting the accuracy of the model. The most problematic is the age group from 0-17 years old.

The following Table 61 contrasts our preliminary hypotheses based on these findings.

Table 61. Preliminary hypotheses and results from the algorithmic impact assessment

Hypotheses	Risk definition	Results from the algorithmic assessment (RISK LEVEL)
MODEL: The combined model might not effectively measure the risk of homelessness in the next 12 months.	Some sub-groups with a high risk of becoming homeless might tend to not use inpatient mental health and ER4 systems, neither to go in jail. Used data/predictors might not be accurately representing the degree and duration of the impairment.	**LOW:** only indirect evidence of limited risk identification was found in the cases of race, gender and veterans. Discrepancies across these group rates were low.
MODEL: People suffering chronic homeless -meaning people suffering from disability (mental or physical)- might be over-protected by the system	People suffering chronic homeless might be over-protected by the system	**LOW:** the algorithm tends to provide high risk of homelessness to people with disability (requirement for chronic homelessness) and also to other outcomes.

RACE: Misrepresentation of Black/African American	Due to the correlation of the information on the race attribute with other factors, such as the amount of information about health for different race groups.	**LOW**: no large discrepancies between races pointing out towards discrimination were found.
GENDER: Lack of female data	The risk assessment is less accurate for the minority group (women) due to a lack of information about them.	**MEDIUM**: overall, women seem to be slightly under-protected by the algorithm.
AGE: Age might become a risk factor in terms of vulnerability	The age attribute may become a basis for false positives and negatives, for instance in the case of elderly people.	**LOW**: No large discrepancies between age groups are found. Rates are in line with age groups under more risk of homelessness.
GENDER/HEALTH: Higher prevalence for certain mental health disorders for women than for men	The algorithm might have been trained based on a clear predominance of data about men, this imbalance could not be correctly captured by the model and predicted by the algorithm.	*Report to follow*
RACE/DISABILITY: Data used to identify disable people could foster bias against Black/African American and low income people	Information and records provided by the Medicare-funded services are used to assess health interactions concerning mental health and other health variables	*Report to follow*
AGE/HEALTH: Since mental health problems seem to increase with age as a risk factor and they importantly determine both the model and the outcomes of algorithmic decisions, the risk for younger segments could be underestimated.	The higher prevalence of psychological and drug abuse issues at a young age and economic and mental problems in adults have also implications in terms of the balance between mental health and substance abuse problems.	*Report to follow*

Source: Own elaboration.

A further analysis will focus on contrasting differences between some groups through an intersectional analysis. Although no significant evidence of discrimination was found, this supplementary analysis is expected to provide further information that may be useful for the system implementation and use.

8.3 Desirability analysis

The Allegheny system is based on a prioritization algorithm that is utilized to organize a ranking through a sorting process. This process is mainly based on data provided by users calling to the DHS. In this regard, we should first consider that historical bias or even realities must sometimes be "excluded" from the training dataset and the model design. This may be done to avoid discrimination, affect the integrity of individuals or maintain ethical standards that go beyond the law (e.g. for instance in the case of parity policies favoring women, which are not legally binding).

Indeed, the **relationship that individuals have with public services can vary from individual to individual** and from social group to social group. For instance, some people might be afraid to go to the doctor and prefer not to go when sick. Others might be less informed about public services and not use them. Others, however, might fear going to the police to report something that has happened to them (for example, irregular immigrants). If an algorithm is aimed at yielding predictions based on these individuals' interaction with the public services, they will be wrongly assessed. This is exactly the type of issue uncovered by Obermeyer et al. (2019). This study found that an algorithm used in public hospitals to determine the level of priority attributed to patients with complex health needs wrongly assigned a higher priority to white people as opposed to black people, even when black people are significantly sicker. This is also due to the fact that the algorithm predicted health care costs rather than illness, but unequal access to healthcare along with a lack of trust in the healthcare system meant that black people were spending less than white people on healthcare (Obermeyer et al., 2019).

This type of unforeseen circumstances affecting the objectivity of the algorithm are to be taken into account when designing it and choosing its proxies. **The desirability of the algorithm also lies in its capacity to adapt to the society in which it operates.**

Furthermore, as it has already been pointed out, the predetermined criteria used by prioritization algorithms to rank individuals are often unavailable to the public, which harms the **accountability of the system** (Diakopoulos, 2015). The higher the impact the algorithm has on individuals, the more important accountability becomes. In the case at hand, the algorithm decides on who is given housing, meaning that those who are not remain in precarious conditions. In this context, it is essential for the algorithm to be transparent to those whom it has an impact on. That means explaining the algorithm; what it does, how it does it, what its impact on people is.

The **desirability of the algorithm may also be differential**. Depending on the protected group we discussed before, the algorithm might be more or less desirable, due to the differential impact we found. At first glance, the social scenario in Allegheny does not seem to be highly problematic either in terms of the variables leading to homelessness or when determining the levels and forms of interaction of different groups of citizens with the administration. Firstly, according to the PIT count, Allegheny has a low rate of homeless people if compared with the national average. Secondly, housing availability is high (64.9% of occupation rate) and poverty rates are low (11.2%) also if we consider the national averages. In spite of all of this, the system must be able to capture qualitative differences across groups and other several casuistic, such as the percentage of people with health insurance, mental illness, housing prices, and drug abuse.

Moreover, desirability must be addressed from the **perspective of resource allocation**, which also conditions the level of tolerable false positives and negatives. The analysis of the system takes into consideration that the algorithm is aimed at assigning a risk level for an adverse outcome, which is to say, the results that can worsen a situation that is already taking a toll on a given individual. In a context in which resources are scarce, this constitutes an ethical issue in itself, since the actual system could be used to limit the possibilities of receiving help in a context of high vulnerability. In this line, the system's development was based on the following normative principles:

> "Ideally we would like **supportive housing to be prioritized for those people who are at highest risk of harm associated with homelessness**; **and where there is heightened risk that they will continue to be homeless if they are not supported into housing**. In other words we want to identify people for whom the largest harms avoided as a result of being provided with services." (DHS-Allegheny County, 2019).

The DHS plans to replace its existing actuarial tool with a Predictive Risk Model (PRM) for different reasons, which are grouped and summarized in Table 62 below.

Table 62. Summary desirability assessment

Policy goals	Explicit reasons behind the development and implementation of the system	Desirability considerations
Automation in order to reduce stress	Minimizing risks of traumatization of clients derived from the use of the current actuarial system by the DHS, which requires them to explain their situation	Automation may harm the transparency and accountability of the system if not applied together with a solid communication strategy
Data minimization	Applying the principle of data minimization by reducing self-reported personal data already in the hands of the DHS	This policy goal is aligned with privacy and data protection principles but it should be accompanied by a strategy for

		registering the automatic decision process
Efficiency	Avoiding adverse outcomes of the current tool, for instance by reducing work time in completing assessments	The accuracy of the model should be constantly monitored.

Source: Own elaboration.

It should be noted that the audited system has not been implemented at the time of completion of this audit report (March 2020). For this reason, it has not been possible to analyze aspects related to the desirability and acceptability of the real implementation of the system within the County procedures in this report. Nor has it been possible to examine the levels of satisfaction, training and confidence in the system of the staff interacting with it. Notwithstanding, the collaboration of the County has made it possible to replace these interviews with an **online interview with Andy Halfhill**, Manager of Homeless/Housing Analytics, Office of Analytics, Technology and Planning of the Allegheny County Department of Human Services. This has made it possible to complete the information in this report about the expected use of the system, its design and its introduction to the staff who will interact with it (mostly explained in section 1), as well as to refine its conclusions and recommendations.

According to this interview, the system may **help in terms of reducing stigma**, fulfilling the above goals. The Allegheny staff is also satisfied with the potential of the system for making **assessments faster**. However, the **accuracy of the algorithm** is still in the process of being improved. In this regard, as already mentioned, this audit did not find major concerns in terms of differential treatment based on protected attributes nor relevant evidence of inaccuracy.

It is important to stress that, depending on the aim and policy framework behind the system (maximize detection of people at risk and/or matching available resource allocation), the system can be configured to be more sensitive if we lower the risk prediction threshold. In this regard, we should differentiate between the allocation of risks to the individual derived from the application of the predictive model and the effective allocation of resources related to the provision of specific programs or services once the system is being implemented. While improving the accuracy and equity of the model is crucial for its desirability, the scope of the housing policy developed by Allegheny County requires a separate assessment and consideration.

Even though the scope of housing policy is not part of this audit it should be noted that the condition of chronic homelessness leads to the possible assignation of permanent housing. Groups of people included in this category should, therefore, be provided with a high-risk

score by the system. As mentioned in the Allegheny County report of 2019, not all of these people can be finally engaged in the permanent housing program, even though they are of high priority. This seems to be the case since the algorithm is giving higher risk to variables related to this condition such as disability. However, since the algorithm also processes other several predictors, highly prioritized individuals may not always be those with a chronic condition. Moreover, the conditions for accessing permanent housing in Allegheny are stricter than those defined by the HUD, since they involve both having a disability plus other vulnerable circumstances, such as "have been living in a place not meant for human habitation, in an emergency shelter, or in a safe haven for at least 12 months either continuously or cumulatively over a period of at least 4 occasions in the last 3 years".
Moreover, **the system may fail in this regard, by providing lower scores to people under this category**. Given this combination of facts, this aspect should be one of the focuses of future monitoring, mostly taking into account the limited number of beds available for people who are candidates for permanent housing. As also already identified during the internal validation of the system, this factor could be problematic, particularly concerning the capacity to identify the degree and duration of the impairment should be assessed.

8.4 Recommendations

Taking into account the above-described results concerning algorithmic accuracy and bias, and the analysis of these results from the policy and desirability standpoints, we will propose a set of recommendations for each of the relevant addressed axes. These recommendations will cover four main segments, concerning the current model, the mechanisms aimed at addressing the limited accuracy of the system to measure the risk of MH inpatient, the ways of addressing potential algorithmic discrimination, and other technical recommendations in case of remodeling or updates of the current model.

8.4.1 Overall accuracy

We consider the **model to be accurate**, as using off-the-shelf tools with some parameter tuning did not yield a better model than this. We did not find any obvious improvements to be done to the modeling part.

However, we did note that **the 0-17 age group is heavily underrepresented** in the training data. This means that in our opinion the combined risk score or other metrics should not be computed for this group, as the accuracy of the model for that group cannot be properly evaluated. This group should be handled separately by a business rule, not by this model.

8.4.2 Algorithmic discrimination

In our analysis we looked for algorithmic discrimination as defined in section 2 above, in terms of whether the algorithm (combined risk scores) introduced a large disadvantage for an already disadvantaged group. **We did not find this to be the case**, however we recommend the following:

I. DISABILITY: We did observe that **people with disabilities are positively discriminated by the algorithm**, which we do not flag as a reason for concern.

II. GENDER: While we did not observe large differences in DI/DT or FNR between men and women, future assessments of the model should **verify if women are more likely to be under-protected than men**, as we did notice a small difference in FNR which could become a disadvantage for them. We recommend to closely monitor the behavior of the overall allocation system with respect to men and women.

III. AGE: We did **not observe large differences of DI/DT or FNR** between age groups that were not explainable by the inherent differences between those groups.

IV. RACE: We **did not observe large differences of DI/DT or FNR between races** that were not explainable by the inherent differences between those groups, or by the fact that the "other" race group is smaller than the other groups and thus statistics on it may exhibit more variability.

V. VETERAN STATUS: We observe **some inconsistencies in FNR** in which veterans are more underprotected in one case (Single) and less underprotected in another case (Family), which means that probably this group should be handled specifically by a business rule.

VI. PROPERTY based discrimination**: No correlational evidence of socioeconomic discrimination has been found**, trough proxies such as race. Still, given the sensitivity of the system it is recommended to monitor this variable in the future by using direct information about income or employment.

8.4.3. Future re-modeling

Homelessness as other social phenomena is expected to experience changes in the future in terms of the composition of the homeless population, the drives for homelessness, its consequences, and so on. We recommend remodeling at least once a year with new data, maintaining some consistency with the methodologies used before, but also introducing new methodological elements if they are found to produce good results.

We also recommend running an analysis of algorithmic fairness in terms of DI/DT and FNR over different groups and combinations of groups, understanding that small differences are inevitable, but looking for large differences that can be addressed by business rules on top of the main model.

Finally, we recommend documenting carefully the training dataset and the modeling process through model cards, and making this documentation public, allowing researchers access to micro-data that allows them, under confidentiality-guaranteeing, research-only agreement, to examine the models closely.

8.4.4 Explainability

By taking advantage of the above-documented capabilities and limitations of the model, as well as the information about the model documented over time, it is recommended to establish a clear communication strategy towards users. All of this information should be made available to clients in a friendly and accessible manner, taking into account also their vulnerable condition. Follow-ups about possible changes in the model should also be integrated into these communication tools.

8.4.5 Desirability

As explained in section 8.3., the predictive system for risk of homelessness audited in this report has not yet been launched by Allegheny County at the time of completion of this report. Notwithstanding, some prior recommendations can be made in relation to an improvement in the desirability of the system:

→ Formal training:

- In order to guarantee the correct preparation of the workers who interact with the model (directly and indirectly), it is recommended to carry out continuous training that allows them to replace their previous dynamics with those of interaction with the risk prediction system, and to incorporate important aspects of the technical functioning of the algorithmic system, its scope and limitations into their professional evaluation.

→ Satisfaction with the system:

- It is recommended to collect data on satisfaction with the system, both from workers who interact directly with it, and from those who receive the priority list where their results are integrated. This would allow to evaluate in a more comprehensive way the desirability and adequacy of the system used, with respect to the previous prioritization system.
- It is recommended to collect data on the level of risk calculated by the system, and the final decision taken by the professional, so as to make possible adjustments to the model in the future".

→ Unified use of the system:

- It is recommended that the integration of the system and its future improvements tend towards a unified use of the risk prediction system by all the DHS staff.

→ Continuous improvement of the system:

- In accordance with the above recommendations, it is also suggested that the system be improved continuously, in relation to the weaknesses detected by the staff and also as part of these audit results (mostly concerning veterans and women).

→ Assess the housing policy of the County in relation to the risk thresholds defined for the algorithm:

- One of the issues identified in this study is that the resources available to the county to combat the street situation are significantly limited. Therefore, while the system may be effective in allocating these resources, they remain scarce. From the algorithmic standpoint, this means that if resources increase, the model could be adjusted or its implementation protocols may be modified.

References

Akthar, R., Winsborough, L., Darko and Chamorro-Premuzic, T. (2019). "Identifying and managing talent in the age of artificial intelligence". In: Fred Oswald, Tara S. Behrend, Lori Foster, *Workforce Readiness and the Future of Work*. New York: Routledge.

Allegheny County (2018). *Allegheny County 2017 Point-in-Time Homelessness Data*. Allegheny: Allegheny County.

Allegheny County (2019). *Allegheny County 2018 Point-in-Time Homelessness Data*. Allegheny: Allegheny County.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias. ProPublica*. Retrieved from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Baeza-Yates, R. (2018). "Bias on the web", *Communications of ACM,* 61 (6), 54-61.

Barabas, C., Virza, M., Dinakar, K., Ito, J., and Zittrain, J. (2018). "Interventions over predictions: Reframing the ethical debate for actuarial risk assessment", *Conference on Fairness, Accountability and Transparency*, 62–76.

Barocas, S. and Hardt, M. (2017). "Fairness in Machine Learning", *Conference on Neural Information Processing Systems, Long Beach, CA, 2017*).

Barocas, S. and Selbst, A. (2016). "Big Data& Disparate Impact", *California Law Review.* Available at: http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf.

Barocas, S., Hardt, M., & Narayanan, A. (2018). *Fairness and machine learning*. Available at : https://fairmlbook.org/pdf/fairmlbook.pdf.

Bartošovič, I. (2016). *Some aspects of the health status of homeless people*. St. Elisabeth: University of Health and Social Sciences.

Bear, J.S., Ginzler, J.A., Peterson, P.L. (2003). DSM-IV alcohol and substance abuse and dependence in homeless youth, *J Stud Alcohol, 64*(1): 5-14. doi: 10.15288/jsa.2003.64.5.

Bender, K., Brown, S. M., Thompson, S. J., Ferguson, K. M., & Langenderfer, L. (2015). Multiple Victimizations Before and After Leaving Home Associated With PTSD, Depression, and Substance Use Disorder Among Homeless Youth, *Child Maltreatment*, 20(2), 115–124. doi: 10.1177/1077559514562859.

Binns, R. (2018). Algorithmic Accountability and Public Reason, *Philos. Technol*., 31 (4), 543-556.

Bolukbasi, Tolga; Kai-Wei Chang, James Y; Zou, Venkatesh Saligrama, and Adam TKalai, (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings, *Advances in Neural Information Processing Systems*: 4349–4357.

Boyd, A., Van de Velde, S., Vilagut, G., de Graaf, R., O'Neill, S., Florescu, S., Alonso, J., Kovess-Masfety, V. (2015). Gender differences in mental disorders and suicidality in Europe: Results from a large cross-sectional population-based study, *Journal of Affective Disorders*, 173, 245-254.

Bradley, C., McGowan, J. & Michelson (2018). How Does Homelessness Affect Parenting Behaviour? A Systematic Critical Review and Thematic Synthesis of Qualitative Research, *D. Clin Child Fam Psychol Rev*, 21 (94). https://doi.org/10.1007/s10567-017-0244-3.

Britton, L. and Pilnik, L. (2018). "Preventing Homelessness for System-Involved Youth", *Juv Fam Court J*, 69, 19-33. doi:10.1111/jfcj.12107.

Brown RT, Goodman L, Guzman D, Tieu L, Ponath C, Kushel MB. (2016). Pathways to Homelessness among Older Homeless Adults: Results from the HOPE HOME Study, *PLoS ONE,* 11(5): e0155065. https://doi.org/10.1371/journal.pone.0155065

Brown, R. T., & Steinman, M. A. (2013). Characteristics of Emergency Department Visits by Older Versus Younger Homeless Adults in the United States, *American Journal of Public Health*, 103(6), 1046–1051. doi:10.2105/ajph.2012.301006.

Byrne, Thomas, Ellen A. Munley, Jamison D. Fargo, Ann E. Montgomery, and Dennis P. Culhane. (2012). "New Perspectives on Community-Level Determinants of Homelessness," *Journal of Urban Affairs,* 35 (5): 607–625.

Carter Narendorf, Sarah. (2017). Intersection of homelessness and mental health: A mixed methods study of young adults who accessed psychiatric emergency services, *Children and Youth Services Review,* 81, 54-62, doi: 10.1016/j.childyouth.2017.07.024.

Castillo (2019). Fairness and Transparency in Ranking, *ACM SIGIR Forum*, 52 (1): 64- 71.

Castillo, C. (2018). Algorithmic Discrimination. Assessing the impact of machine intelligence on human behaviour: an interdisciplinary endeavour, *Proceedings of HUMAINT Workshop*. Available at: https://arxiv.org/pdf/1806.03192.pdf.

Chen, I. Y. ; Peter Szolovits, and Marzyeh Ghassemi. (2019). Can AI Help Reduce Disparities in General Medical and Mental Health Care?, *AMA J Ethics.,* 21(2): 167-179.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments, *arXiv:1610.07524*.

Chouldechova, A.; D. Benavides-Prado, O. Fialko, and R. Vaithianathan, (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency,* Friedler and C. Wilson, Eds., 81, New York, NY, USA: PMLR, 23–24 Feb 2018: 134–148.

City of Pittsburgh, Pennsylvania Code of Ordinances (2019). Codified through Ordinance No. 33-2019, October 22.

Clark, R.E., Weinreb, L. Flahive, J. M. and W. Seifert. R.W. (2019). Homelessness Contributes To Pregnancy Complications, *Health Affairs,* 38(1): 139-146.

Community Services Committee (2007). City of Melbourne Homelessness Framework, 2007 - 2009.

Cowgill, Bo (2019). Bias and productivity in humans and machines, *Upjohn Institute Working Paper, No. 19-309, W.E. Upjohn Institute for Employment Research,Kalamazoo*, MI biases-1508252717.

Craig, T., & Hodson, S. (1998). Homeless youth in London: I. Childhood antecedents and psychiatric disorder, *Psychological Medicine*, 28: 1379–1388. doi 1017/S0033291798007351

Culhane, Dennis P. ; Stephen Metraux, Jung Min Park, Maryanne Schretzman & Jesse Valente (2007). Testing a typology of family homelessness based on patterns of public shelter utilization in four U.S. jurisdictions: Implications for policy and program planning, *Housing Policy Debate*, 18:1, 1-28, doi: 10.1080/10511482.2007.9521591

Danks, David and Alex John London (2017). Algorithmic bias in autonomous systems, *Proceedings of the 26th International Joint Conference on Artificial Intelligence.AAAI Press, 2017:* 4691–4697.

Diakopoulos, Nicholas (2015). Algorithmic Accountability, *Digital Journalism,* 3(3): 400-403.

Dwork, C. and Ilvento, C. (2018b). Individual fairness under composition, *FATML.*

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness, *Proceedings of the 3rd innovations in theoretical computer science conference*, 214-226. ACM.

Eubanks, V. (2018). *Automating Inequality*. New York: St. Martin's Press.

Feldmann, J., & Middleman, A. B. (2003). Homeless adolescents: Common clinical concerns. *Seminars in Pediatric Infectious Diseases, 14*(1), 6–11. doi https://doi.org/10.1053/spid.2003.127211

Fernando, Suman, Keating, Frank, eds. (2008). *Mental Health in a Multi-Ethnic Society. A Multidisciplinary Handbook*. London: Routledge.

Fischer S.N., Shinn M., Shrout P., Tsemberis S. (2008). Homelessness, mental illness, and criminal activity: examining patterns over time, *Am J Community Psychol, 42*(3-4): 251-65. doi: 10.1007/s10464-008-9210-z.

Foulds, James R. and Shimei Pan. (2018). An Intersectional Definition of Fairness, *CoRRabs/1807.08362*.

Gillespie, T. 2012. "The Relevance of Algorithms". In: *Media Technologies*, edited by Gillespie, Tarleton , Boczkowski, Pablo and Foot, Kirsten. Cambridge, MA: MIT Press.

Glendening, Z., & Shinn, M. (2017). Risk Models for Returns to Housing Instability Among Families Experiencing Homelessness, *Cityscape (Washington, D.C.), 19*(3), 309–330.

Glynn, Chris, and Emily B. Fox. (2017). *Dynamics of Homelessness in Urban America*. Ithaca, NY: Cornell University, arXiv. http://arxiv.org/abs/1707.09380.

Goldhaber-Fiebert, J. D. (2019). "Impact evaluation of a predictive risk modeling tool for allegheny countys child welfare office," [Online]. Available at: https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-6.pdf.

Gonzalez, J.R., Jetelina, K.K., Roberts, M. e Am J Crim Just (2018). Criminal justice system involvement among homeless adults, *Am J Crim Just:* 18, 1-9.

Green, J. M., Ennett, S. T., & Ringwalt, C. L. (1997). Substance use among runaway and homeless youth in three national samples, *American Journal of Public Health, 87*: 229-235.

Gubits, Daniel; Shinn, Marybeth; Wood, Michelle; Bell, Stephen; Dastrup, Samuel; Solari, Claudia; Brown, Scott; McInnis, Debi; McCall, Tom; Kattel, Utsav, (2016). *Family Options Study: 3-Year Impacts of Housing and Services Interventions for Homeless Families*. U.S. Department of Housing and Urban Development.

Hajian, S., Bonchi, F., y Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining, *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*: 2125-2126.

Heidari, H., Ferrari, C., Gummadi, K., and Krause, A. (2018). Fairness behind a veil of ignorance: a welfare analysis for automated decision making, in *Advances in Neural Information Processing Systems 31*, eds S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Montreal, QC: Curran Associates, Inc), 1265–1276.

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 600). ACM.

HUD (2018). The 2018 Annual Homeless Assessment Report (AHAR) to Congress. The U.S. Department of Housing and Urban Development.

Humes, Karen R. ; Nicholas A. Jones; Roberto R. Ramirez (March 2011). *Overview of Race and Hispanic Origin: 2010*. United States Census Bureau.

James, DJ. (2004). *Bureau of Justice statistics special report: profile of jail inmates, 2002.* Washington, DC: U.S. Department of Justice.

Johnson,Robin; Ray Middleton, Suzanne Quinney, John Conolly, Joe Finn (2017). *Cross-Cultural Dialogues on Homelessness: From Pretreatment Strategies to Psychologically Informed Environments*. MI: Loving Healing Press.

Keeshin, B.R. and Campbell, K.(2011). Screening homeless youth for histories of abuse: Prevalence, enduring effects, and interest in treatment, *Child Abuse & Neglect,* 35: 6, 401-407. https://doi.org/10.1016/j.chiabu.2011.01.015.

Kessler, R.C., McGonagle, K.A. and S. Zhao S. et al. (1994). Lifetime and 12 month prevalence of DSM-111-R psychiatric disorders in the United States, *Archives of General Psychiatry*, 51, 8-19.

Kim, Michael P, Reingold, Omer, and Rothblum, Guy N. (2018). Fairness through computationally-bounded awareness; *arXiv preprint arXiv:1803.03239*.

Kithulgoda, Chamari I., Emily Kulick, Abigail Horn, Erin Dalton, Sophie Joyce and Rhema Vaithianathan. (2017). *Implementing a Predictive Risk Model to Prioritize Homeless Services: The Allegheny Homelessness Tool*. Center for Social Data Analytics. Auckland University of Technology. Available at:

https://csda.aut.ac.nz/__data/assets/pdf_file/0005/309146/The-Allegheny-Homelessness-Tool-by-CSDA-and-ACDHSl-BloombergPoster-ForPrint.pdf.

Kroll, Joshua A, Joanna Huey, Solon Barocas, Edward Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu (2017). Accountable Algorithms, *University of Pennsylvania Law Review* (165).

Kyung Lee, Min (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management, *Big Data & Society*, 5, 1.

Langan, N. P., & Pelissier, B. M.M. (2001). Gender differences among prisoners in drug treatment, *Journal of Substance Abuse, 13*(3), 291–301. doi: 10.1016/S0899-3289(01)00083-9.

Lippert-Rasmussen, Kasper (2013). *Born Free and Equal? A Philosophical Inquiry Into the Nature of Discrimination*. Oxford: Oxford University Press.

Martijn, Claudine, Louise Sharpe, (2006). Pathways to youth homelessness, *Social Science & Medicine*, 62 (1), 1-12, ISSN 0277-9536, doi: 10.1016/j.socscimed.2005.05.007.

McKenzie, Raub (2018). Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices, *71 Ark. L. Rev.* 529.

Mitchell, J.N., Clark, C. & Guenther, C.C. Community Ment Health J (2017). The Impact of Criminal Justice Involvement and Housing Outcomes Among Homeless Persons with Co-occurring Disorders, *C.C. Community Ment Health J,* 53: 901. doi 10.1007/s10597-017-0089-8

Narayanan, A. (2018). Tutorial: 21 definitions of fairness and their politics [Abstract and video] Conference on Fairness, Accountability, and Transparency, NYC Feb 23.

Narendorf, S. (2017). Intersection of homelessness and mental health: A mixed methods study of young adults who accessed psychiatric emergency services, *Children and Youth Services Review, 81*(2017): 54-62. doi: 10.1016/j.childyouth.2017.07.024.

National Health Care for the Homeless Council (2017). *Annual Report.*

Ní Cheallaigh C, Cullivan S, Sears J, et al. (2017). Usage of unscheduled hospital care by homeless individuals in Dublin, Ireland: a cross-sectional study, *BMJ Open* 2017;7:e016420. doi: 10.1136/bmjopen-2017-016420

Nisar, Hiren; Mallory Vachon, Charles and Horseman, Jim (2019). *Market Predictors of Homelessness: How Housing and Community Factors Shape Homelessness Rates Within Continuums of Care*. USA: DHU.

Nissenbaum, Helen (2001). How computer systems embody values, *Computer*, 34 (3): 120–119.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations, *Science*, 366(6464), 447-453.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Broadway Books.

Pasquale, Frank (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Remster, B. (2019). A Life Course Analysis of Homeless Shelter Use among the Formerly Incarcerated, *Justice Quarterly*, 36 (3), 437-465, doi 10.1080/07418825.2017.1401653

Riden, M. (2011). *From the Streets to Stability: A study of youth homelessness in the District of Columbia*. Columbia: DCAYA.

Rosenthal, D., Mallet, S. and Myers, P. (2006). Why do homeless young people leave home?, *Australian and New Zealand Journal of Public Health, 30*(3): 281-285. doi 10.1111/j.1467-842X.2006.tb00872.x.

Ryan, A.M., Gee, G.C., & Laflamme, D.F. (2006). The Association between Self-Reported Discrimination, Physical Health and Blood Pressure: Findings from African Americans, Black Immigrants, and Latino Immigrants in New Hampshire, *Journal of Health Care for the Poor and Underserved 17*(2), 116-132. doi:10.1353/hpu.2006.0079.

SAMHSA (2002). National Survey on Drug Use and Health (NSDUH-2002). Retrieved from: https://www.datafiles.samhsa.gov/study/national-survey-drug-use-and-health-nsduh-2002-nid13623

Schütz, Christian, Fiona Choi, Michael Jae Song, Christiane Wesarg, Kathy Li & Michael Krausz (2019). Living With Dual Diagnosis and Homelessness: Marginalized Within a Marginalized Group, *Journal of Dual Diagnosis*, 15:2, 88-94, doi 10.1080/15504263.2019.1579948

Shah, M. F., Liu, Q. , Mark Eddy, J. , Barkan, S. , Marshall, D. , Mancuso, D. , Lucenko, B. and Huber, A. (2017), Predicting Homelessness among Emerging Adults Aging Out of Foster Care, *Am J Community Psychol,* 60: 33-43. doi:10.1002/ajcp.12098

Shinn, M., Greer, A. L., Bainbridge, J., Kwon, J., & Zuiderveen, S. (2013). Efficient targeting of homelessness prevention services for families. *American journal of public health,* 103 Suppl 2(Suppl 2), 324–S330. doi:10.2105/AJPH.2013.301468

Shinn, M., Weitzman, B. C., Stojanovic, D., Knickman, J. R., Jiménez, L., Duchon, L., Krantz, D. H. (1998). Predictors of homelessness among families in New York City: from shelter request to housing stability, *American journal of public health,* 88(11), 1651–1657. doi:10.2105/ajph.88.11.1651

Shinn, Marybeth; Andrew L. Greer, Jay Bainbridge, Jonathan Kwon, Sara Zuiderveen, “Efficient Targeting of Homelessness Prevention Services for Families”, *American Journal of Public Health,* 103(2), 324-330. doi 10.2105/AJPH.2013.301468.

Smith, Nancy; Zaire Dinzey Flores, Jeffrey Lin, John M. (2005). *Understanding family homelessness in New York city. An In-Depth Study of Families’ Experiences Before and After Shelter*. New York: Vera Institute of Justice.

Solorio, M. R., Milburn, N. G., Andersen, R. M., Trifskin, S., & Gelberg, L. (2006). Health care service use among vulnerable adolescents, *Vulnerable Children and Youth Studies*,1(3), 205–220. doi: 10.1080/17450120600973437.

Speak, Suzanne (2019) The State of Homeless in Developing Countries. Available at: https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2019/05/SPEAK_Suzanne_Paper.pdf

Speicher, Till et al. (2018). A Unified Approach to Quantifying Algorithmic Unfairness, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*.

Steinhaus, Douglas A; Harley, Debra A; Rogers, Jackie (2004). Homelessness and People with Affective Disorders and other Mental Illnesses, *Journal of Applied Rehabilitation Counseling*; 35 (1), 36-40.

Suresh, Harini and John V Guttag. (2019). A Framework for Understanding Unintended Consequences of Machine Learning, *arXiv preprint* arXiv:1901.10002.

The Coalition for Juvenile Justice (CJJ)(2016). *Policy Brief on Youth Homelessness and Juvenile Justice: Opportunities for Collaboration and Impact.* Available at: http://www.juvjustice.org/sites/default/files/resource-files/policy%20brief_FINAL.compressed.pdf.

The National Coalition for the Homeless and the National Law Center on Homelessness and Poverty. *A Dream Denied: The Criminalization of Homelessness in U.S. Cities*. https://www.nationalhomeless.org/publications/crimreport/report.pdf.

The US Department of Housing and Urban Development (2018). *The 2018 Annual Homeless Assessment Report (AHAR) to Congress*.

Tibshirani, Robert (1996). Regression Shrinkage and Selection via the lasso, *Journal of the Royal Statistical Society, 58*(1): 267–88. JSTOR 2346178

Toros, Halil and Flaming, Daniel, (2018). Prioritizing Homeless Assistance Using Predictive Algorithms: An Evidence-Based Approach. *Cityscape, 20*(1).

U.S. Department of Health and Human Services. (2014): *Substance Abuse and Mental Health Services Administration* (US).

Understanding Homelessness (2019). http://www.understandhomelessness.com.

United States Census Bureau, (2018). *QuickFacts.* Allegheny County, Pennsylvania. Available at: https://www.census.gov/quickfacts/fact/table/alleghenycountypennsylvania/PST045218.

United States Interagency Council on Homelessness (2010). *Opening doors: Federal strategic plan to prevent and end homelessness*. Available at: https://www.usich.gov/resources/uploads/asset_library/USICH_OpeningDoors_Amendment2015_FINAL.pdf.

Vaithianathan, R.; E. Putnam-Hornstein, N. Jiang, P. Nand, and T. Maloney. (2017). Developing predictive models to support child maltreatment hotline screening decisions: Allegheny county methodology and implementation, [Online]. Available at: https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf.

Vedder, Anton; Naudts, Laurens; 2017. Accountability for the Use of Algorithms in a Big Data Environment. *International Review of Law, Computers & Technology, 31*(2): 206 – 224

Walker, S., Valencia, E., Bishop, A., Irons, M., & Gertseva, A. (2018). Developing a Coordinated Youth Housing Stability Program for Juvenile Courts, *Cityscape*, 20(3), 117-138. Available at: https://www.jstor.org/stable/26524875.

Wallace, S. (1965). *Skid Row as a Way of Life.* New York: Harper and Row.

Weiser, S. D., Neilands, T. B., Comfort, M. L., Dilworth, S. E., Cohen, J., Tulsky, J. P., & Riley, E. D. (2009). Gender-specific correlates of incarceration among marginally housed individuals in San Francisco, *American Journal of Public Health*, *99*(8): 1459–1463

Wong, C. F., Clark, L. F., & Marlotte, L. (2016). The impact of specific and complex trauma on the mental health of homeless youth, *Journal of Interpersonal Violence*, *31*(5), 831–854. doi 10.1177/0886260514556770.

Wong, Y. J., Ho, M.-H. R., Wang, S.-Y., & Miller, I. S. K. (2017). Meta-analyses of the relationship between conformity to masculine norms and mental health-related outcomes, *Journal of Counseling Psychology, 64*(1), 80–93. doi: 10.1037/cou0000176

World Health Organization (2002). *Gender and Mental Health*. Available at: https://apps.who.int/iris/bitstream/handle/10665/68884/a85573.pdf?sequence=1&isAllowed=y.

Zerger, S., Strehlow, A. J., & Gundlapalli, A. V. (2008). Homeless Young Adults and Behavioral Health: An Overview, *American Behavioral Scientist,* 51(6), 824–841. doi 10.1177/0002764207311990.