

How to Measure Bias in AI

Internal Interrogation – Compare the ideal target to the actual target for at risk populations

1. **Define the ideal target & the actual target**
 - *What do you want the algorithm to learn?*
2. **Choose comparison groups**
 - *Who is at risk of bias?*
3. **Compare the populations within the training dataset with the populations in which the algorithm is being applied**
 - *How are underserved groups defined?*
 - *What is the fraction of each underserved group?*
4. **Compare algorithm results across comparison groups**
 - *Do patients have the same level of the actual target?*
5. **Measure performance between ideal target and actual target, using a diagnostic chart if possible**

| Algorithm | Ideal Target | Actual Target | Risk of Bias |
|---|--|--|--|
| Care Management Prioritization: Identifying patients for additional services | Health needs, benefit from high-risk care management programs | Total costs of care | High. Less money is spent on Black patients who have the same level of need |
| Emergency Severity Index (ESI): emergency triage | Medical condition needing immediate attention | Nurse-rated acuity, “resources patient is expected to consume” | High. Resource consumption varies by race and insurance for any given acuity |
| 6-Clicks Mobility Score: Decisions about discharge destination | Inability to care for self and live independently at home without help | Physical measures of mobility and daily activities | High. Similar physical mobility scores have larger impact on those lacking income |
| “No-show” prediction: Clinic scheduling | Voluntary no-show to appointment | Any no-show to prior appointment | High. No shows relate to access: barriers are unequally distributed |
| Predicting Disease Onset: Targeting preventative care | New disease onset (e.g., heart failure, kidney failure) | Provider–insurer transaction with ICD code for disease | High. Probability of being coded varies by physician quality, hospital billing, insurance, etc. |
| Kellgren–Lawrence Grade: Osteoarthritis on knee x-rays | Severity of knee osteoarthritis | Severity of osteoarthritis seen by radiologist on knee x-rays | High. Radiologists miss causes of knee pain affecting underserved groups |

When screening for label choice bias, organizations should determine the extent to which the discrepancy between the ideal and actual target is likely to create bias for underserved groups.