Supplementary Table 1 | Examples of machine-learning tasks with fairness issues in medicine and healthcare. We outline machine-learning tasks in several areas of medicine, current or potential issues in fairness, and their associated dataset shift. We further differentiate between non-biological and biological factors that cause population shifts, which we respectively further attribute as being driven by social determinants of health (SD) or by genetic ancestry (ancestry).

Area of medicine	Machine-learning task	Issue in fairness	Type of dataset shift
Clinical lab measurements and electronic medical records	Predicting kidney failure using the eGFR equation	Predicting kidney failure using eGFR equation & race-specific covariate bias kidney function appears better in Black patients, which could delay medication and referrals for precluding kidney failures[1–6].	Population shift (SD)
	Checking for uterine track infection	Race-specific covariate bias assigns lower odds of checking checked for UTI in Black patients, reduces likelihood of scheduling follow-up and referrals[7].	Population shift (SD)
	Predicting osteoporosis using a bone fracture risk calculator	Race-specific covariate place black women at lower risk of osteoporosis, while high-risk patients receive preventative drugs to minimize fractures[8].	Population shift (SD)
	Opioid early-warning system	Changing from ICD-9 to ICD-10 resulted in a large wave of false negatives and a much higher prevalence of opioid-related codes[9].	Concept shift
	Risk prediction	An algorithm that used health costs as a proxy for health needs would predict Black patients as being lower risk than equally-sick White patients[10].	Population shift (SD), Label shift
Genomics	Polygenic risk scores	Variations in linkage disequilibrium structures and minor allele frequencies across ancestral populations contributes to worse performance of genetic polygenic risk models in underrepresented populations[11].	Population shift (ancestry)
	Cancer prognosis	Genomic tests for prostate cancer prognosis, which may have been developed with individuals of primarily European ancestry, may predict perform worse on underrepresented populations[12].	Population shift (SD)
	Response-to-treatment prediction	Cell lines such as the E006AA-hT prostate cancer cell line, misclassified as African American, have been found to carry 92% European ancestry[13]. Such misclassification would mislead models developed on data based on this cell line, as well as healthcare disparities and fairness research.	Label shift
		Ancestry-specific innate immune variants contribute toward higher incidence and mortality of Triple Negative Breast Cancer among individuals of African ancestry[14–18]. AI algorithms developed without the inclusion of ancestry may worse performance on this group of patients.	Population shift (ancestry)
	Response-to-treatment prediction	Black patients are overwhelmingly underrepresented in clinical trials (less than 2% of NCI-funded clinical trials include non-White patients)[19,20]. Application of AI-based methods for biomarker discovery to retrospective clinical trial cohorts may have poor generalization performance on non-White patients.	Sample selection bias

Radiology	Disease segmentation and detection in MRI / CT / chest X-rays / mammography scans	AI algorithms trained on publicly-available radiology images misdiagnose under-served patients at a disproportionate rate compared to the baseline population[21].	Population shift (SD)
		Model leakage of self-reported ethnicity information after controlling for site-specific technical artifacts and potential anatomic differences[22].	Unknown
Pathology	Cancer diagnosis, prognosis, response-to-treatment prediction, mutation prediction from H&E	Genetic variation amongst patients of different ethnicity, ancestry, geographic locations and other environmental factors[23–31] may result in population-specific phenotypes[32] and lead to disparities in diagnostic and prognostic algorithms that use histology[33,34].	Population shift (ancestry)
		Only patients developing symptoms will be biopsied, which produces disparities in patients who will get pathology services due to access to care, leading to dataset imbalance.	Sample selection bias
		H&E stain intensity can predict ethnicity on the cancer genome atlas (TCGA), owing to hospital-specific image-acquisition protocols[35].	Acquisition shift
		Evolution of novel diseases and their comorbidities may bias deployment of current models[36–38].	Open set label shift
	Renal allograft assessment	Taxonomies such as the Banff classification system for renal allograft assessment are updated with new diagnostic criteria every two years[39].	Concept shift
	Predicting tumour origin in cancers of unknown primary	AI algorithms that do not include patient sex may diagnose patients with unlikely and incorrect tumour origins[40].	Population shift (ancestry)
Ophthalmology	Retinopathy grading, risk assessment, and vessel segmentation	Fundus photography images have been demonstrated to not only cardiovascular risk factors, but also traits such as age and gender[41,42]. Phenotypic variations such as melanin concentration and retinal-vessel appearance have also been shown to differ across demographics[43].	Population shift (ancestry)
		AI screening tools for diabetic retinopathy developed in the U.S., may fail to generalize to countries in Southeast Asia due to varied lighting conditions and socio-economic factors of how the screening is performed by nurses[44].	Acquisition shift
		Differing clinical education in training ophthalmologists, as well as intraobserver variability, may cause label bias in training AI algorithms[44,45].	Label shift
Rheumatology	Predicting pain and surgery eligibility	Disparities in how different populations respond to pain, may bias algorithms trained on reported pain score[46].	Population shift (SD)
Dermatology	Skin-lesion classification	ML-based mobile health apps may not have been developed with darker skin types in the train dataset, which may over- or under-diagnosis non-White patients with under-represented Fitzpatrick skin types[47,48].	Sample selection bias
	Biometrics monitoring via wearables	Patients with higher melanin may block green light used by wearable devices for accurately measuring heart rate[49,50].	Population shift

References

1. Vyas, D. A., Eisenstein, L. G. & Jones, D. S. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms (2020).
2. van der Burgh, A. C., Hoorn, E. J. & Chaker, L. Removing race from kidney function estimates. JAMA 325, 2018–2018 (2021).
3. Diao, J.A., Wu, G.J., Taylor, H.A., Tucker, J.K., Powe, N.R., Kohane, I.S. and Manrai, A.K., 2021. Clinical implications of removing race from estimates of kidney function. JAMA, *325*, 184-186 (2021).
4. Eneanya, N.D., Yang, W. and Reese, P.P. Reconsidering the consequences of using race to estimate kidney function. *JAMA*, *322*, 113-114 (2019).
5. Zelnick, L.R., Leca, N., Young, B. and Bansal, N., 2021. Association of the estimated glomerular filtration rate with vs without a coefficient for race with time to eligibility for kidney transplant. *JAMA network open*, *4*, e2034004-e2034004 (2021).
6. Chadban, S.J., Ahn, C., Axelrod, D.A., Foster, B.J., Kasiske, B.L., Kher, V., Kumar, D., Oberbauer, R., Pascual, J., Pilmore, H.L. and Rodrigue, J.R., 2020. KDIGO clinical practice guideline on the evaluation and management of candidates for kidney transplantation. Transplantation, 104, S11-S103 (2020).
7. Vyas, D. A. et al. Challenging the use of race in the vaginal birth after cesarean section calculator. Women's Health Issues 29, 201–204 (2019).
8. Curtis, J. R. et al. Population-based fracture risk assessment and osteoporosis treatment disparities by race and gender. Journal of general internal medicine 24, 956–962 (2009).
9. Heslin, K. C. et al. Trends in opioid-related inpatient stays shifted after the us transitioned to icd-10-cm diagnosis coding in 2015. Medical care 55, 918–923 (2017).
10. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 366, 447–453 (2019).
11. Li, S., Cai, T. & Duan, R. Targeting underrepresented populations in precision medicine: A federated transfer learning approach. Annals of Applied Statistics (2023).
12. Creed, J. H. et al. Commercial gene expression tests for prostate cancer prognosis provide paradoxical estimates of race-specific risk. Cancer Epidemiology, Biomarkers & Prevention 29, 246–253 (2020).
13. Hooker Jr, S.E., Woods-Burnham, L., Bathina, M., Lloyd, S., Gorjala, P., Mitra, R., Nonn, L., Kimbro, K.S. and Kittles, R.A.. Genetic ancestry analysis reveals misclassification of commonly used cancer cell lines. *Cancer Epidemiology, Biomarkers & Prevention*, *28*, 1003-1009 (2019).
14. Zavala, V.A., Bracci, P.M., Carethers, J.M., Carvajal-Carmona, L., Coggins, N.B., Cruz-Correa, M.R., Davis, M., de Smith, A.J., Dutil, J., Figueiredo, J.C. and Fox, R.. Cancer health disparities in racial/ethnic minorities in the United States. *British journal of cancer*, *124*, 315-332 (2021).
15. Martini, R., Delpe, P., Chu, T.R., Arora, K., Lord, B., Verma, A., Bedi, D., Karanam, B., Elhussin, I., Chen, Y. and Gebregzabher, E.. African ancestry–associated gene expression profiles in triple-negative breast cancer underlie altered tumor biology and clinical outcome in women of African descent. *Cancer Discovery*, *12,* 2530-2551 (2022).
16. Martini, R., Newman, L. and Davis, M.. Breast cancer disparities in outcomes; unmasking biological determinants associated with racial and genetic diversity. *Clinical & experimental metastasis*, *39*, 7-14 (2022).
17. Yeyeodu, S.T., Kidd, L.R. and Kimbro, K.S.. Protective innate immune variants in racial/ethnic disparities of breast and prostate cancer. *Cancer immunology research*, *7*, 1384-1389 (2019).
18. Iqbal, J., Ginsburg, O., Rochon, P.A., Sun, P. and Narod, S.A.. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. *Jama*, *313*, 165-173 (2015).
19. Herbst, R.S., Giaccone, G., de Marinis, F., Reinmuth, N., Vergnenegre, A., Barrios, C.H., Morise, M., Felip, E., Andric, Z., Geater, S. and Özgüroğlu, M.. Atezolizumab for first-line treatment of PD-L1–selected patients with NSCLC. *New England Journal of Medicine*, *383*, 1328-1339 (2020).
20. Chen Jr, M.S., Lara, P.N., Dang, J.H., Paterniti, D.A. and Kelly, K., Twenty years post-NIH Revitalization Act: Enhancing minority participation in clinical trials (EMPaCT): Laying the groundwork for improving minority clinical trial accrual: Renewing the case for enhancing minority participation in cancer clinical trials. *Cancer*, *120*, 1091-1096 (2014).
21. Seyyed-Kalantari, L., Zhang, H., McDermott, M., Chen, I. Y. & Ghassemi, M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature medicine 27, 2176–2182 (2021).
22. Gichoya, J. W. et al. Ai recognition of patient race in medical imaging: a modelling study. The Lancet Digital Health (2022).
23. Shi, Y. et al. A prospective, molecular epidemiology study of egfr mutations in asian patients with advanced non–small-cell lung cancer of adenocarcinoma histology (pioneer). Journal of thoracic oncology 9, 154–162 (2014).

24. Gusev, A., Shi, H., Kichaev, G., Pomerantz, M., Li, F., Long, H.W., Ingles, S.A., Kittles, R.A., Strom, S.S., Rybicki, B.A. and Nemesure, B.. Atlas of prostate cancer heritability in European and African-American men pinpoints tissue-specific regulation. *Nature communications 7*, 10979 (2016).
25. Valle, L.F., Nickols, N.G., Hausler, R., Alba, P.R., Anglin-Foote, T., Perez, C., Yamoah, K., Rose, B.S., Kelley, M.J., DuVall, S.L. and Garraway, I.P.. Actionable Genomic Alterations in Prostate Cancer Among Black and White United States Veterans. *The Oncologist* (2023).
26. Conti, D.V., Wang, K., Sheng, X., Bensen, J.T., Hazelett, D.J., Cook, M.B., Ingles, S.A., Kittles, R.A., Strom, S.S., Rybicki, B.A. and Nemesure, B.. Two novel susceptibility loci for prostate cancer in men of African ancestry. *JNCI: Journal of the National Cancer Institute*, *109 (*2017).
27. Mahal, B.A., Alshalalfa, M., Kensler, K.H., Chowdhury-Paulino, I., Kantoff, P., Mucci, L.A., Schaeffer, E.M., Spratt, D., Yamoah, K., Nguyen, P.L. and Rebbeck, T.R.. Racial differences in genomic profiling of prostate cancer. *New England Journal of Medicine*, *383*, 1083-1085 (2020).
28. Jones, J., Mukherjee, A., Karanam, B., Davis, M., Jaynes, J., Reams, R.R., Dean-Colomb, W. and Yates, C.. African Americans with pancreatic ductal adenocarcinoma exhibit gender differences in Kaiso expression. *Cancer letters 380*, 513-522 (2016).
29. Borrell, L.N., Elhawary, J.R., Fuentes-Afflick, E., Witonsky, J., Bhakta, N., Wu, A.H., Bibbins-Domingo, K., Rodríguez-Santana, J.R., Lenoir, M.A., Gavin III, J.R. and Kittles, R.A.. Race and genetic ancestry in medicine—a time for reckoning with racism. *New England Journal of Medicine*, *384*, 474-480 (2021).
30. Hinch, A.G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C.D., Chen, G.K., Wang, K., Buxbaum, S.G., Akylbekova, E.L. and Aldrich, M.C.. The landscape of recombination in African Americans. *Nature*, *476*, 170-175 (2011).
31. Shriver, M.D., Parra, E.J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C., Massac, A., Cameron, N. and Baron, A.. Skin pigmentation, biogeographical ancestry and admixture mapping. *Human genetics*, *112*, 387-399 (2003).
32. Bhargava, H. K. et al. Computationally derived image signature of stromal morphology is prognostic of prostate cancer recurrence following prostatectomy in african american patients. Clinical Cancer Research 26, 1915–1923 (2020).
33. Clarke, M.A., Devesa, S.S., Hammer, A. and Wentzensen, N.. Racial and ethnic differences in hysterectomy-corrected uterine corpus cancer mortality by stage and histologic subtype. *JAMA oncology 8*, 895-903.
34. Kehl, K.L., Lathan, C.S., Johnson, B.E. and Schrag, D.. Race, poverty, and initial implementation of precision medicine for lung cancer. *JNCI: Journal of the National Cancer Institute*, *111*, 431-434 (2019).
35. Howard, F. M. et al. The impact of site-specific digital histology signatures on deep learning model accuracy and bias. Nature communications 12, 1–13 (2021).

36. 23 Yehia, B. R. et al. Association of race with mortality among patients hospitalized with coronavirus disease 2019 (covid-19) at 92 us hospitals. JAMA network open 3, e2018039–e2018039 (2020).
37. 24 Lopez, L., Hart, L. H. & Katz, M. H. Racial and ethnic health disparities related to covid-19. JAMA 325, 719–720 (2021).
38. Garg, S., Balakrishnan, S. and Lipton, Z.C., Domain Adaptation under Open Set Label Shift. In *Advances in Neural Information Processing Systems* (2022).
39. Loupy, A., Mengel, M. & Haas, M. Thirty years of the international banff classification for allograft pathology: the past, present, and future of kidney transplant diagnostics. Kidney International (2022).
40. Lu, M.Y., Chen, T.Y., Williamson, D.F., Zhao, M., Shady, M., Lipkova, J. and Mahmood, F., 2021. AI-based pathology predicts origins for cancers of unknown primary. *Nature 594*, 106-110 (2022).
41. Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering 2, 158–164 (2018).
42. Babenko, B. et al. Detection of signs of disease in external photographs of the eyes via deep learning. Nature biomedical engineering 1–14 (2022).
43. Burlina, P., Joshi, N., Paul, W., Pacheco, K. D. & Bressler, N. M. Addressing artificial intelligence bias in retinal diagnostics. Translational Vision Science & Technology 10, 13–13 (2021).
44. Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P. and Vardoulakis, L.M.. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, 12 (2020).
45. Ruamviboonsuk, P., Teerasuwanajak, K., Tiensuwan, M., Yuttitham, K. and Thai Screening for Diabetic Retinopathy Study Group. Interobserver agreement in the interpretation of single-field digital fundus images for diabetic retinopathy screening. Ophthalmology, *113*, 826-832 (2006).
46. Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S. & Obermeyer, Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. Nature Medicine 27, 136–140 (2021).

47. Daneshjou, R., Vodrahalli, K., Novoa, R.A., Jenkins, M., Liang, W., Rotemberg, V., Ko, J., Swetter, S.M., Bailey, E.E., Gevaert, O. and Mukherjee, P.. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science advances*, *8* (2022).
48. Wen, D. et al. Characteristics of publicly available skin cancer image datasets: a systematic review. The Lancet Digital Health (2021).
49. Shcherbina, A., Mattsson, C.M., Waggott, D., Salisbury, H., Christle, J.W., Hastie, T., Wheeler, M.T. and Ashley, E.A.. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of personalized medicine*, *7*, 3 (2017).
50. Costin, G.E. and Hearing, V.J.. Human skin pigmentation: melanocytes modulate skin color in response to stress. *The FASEB journal*, *21*, 976-994 (2007).