

## APPENDIX

### Supplementary Methods

#### Timing of Demographic Information

For patients in *Michigan Medicine*, demographic information was based on the latest hospital visit within the time frame of this study (by selecting the latest encounter, we skew the Michigan Medicine cohort to be slightly older; even so it is younger than *MGI Recruited* and *MGI Enrolled* as discussed in **Results**). For *MGI Recruited* and *MGI Enrolled*, we considered the hospital visit within up to 30 days around the recorded date when the patient was asked to participate in MGI.

#### Relative Participation Rates of Demographic Subgroups

To gain a more holistic view of the overall differences between *Michigan Medicine* and *MGI Enrolled*, we calculated the relative participation rate of each subgroup (e.g., Black or African American patients) as the ratio of (1) that subgroup's participation rate and (2) a baseline participation rate. We considered two baseline participation rates: those not in the target subgroup (e.g., non-Black or African American patients), and the participation rate of the entire cohort. For racial subgroups, we additionally considered a third baseline of White patients' participation rate. For the disadvantage index, we binarized the values with respect to the median of *Michigan Medicine* (for race and ethnicity subgroup analyses, the median values were for the corresponding subgroup in *Michigan Medicine*). A ratio of 1 corresponds to no difference in participation rates. A ratio  $<1$  or  $>1$  suggests that those in the target subgroup participate at a relatively *lower* or *higher* rate compared to the baseline.

#### Measuring Effect of Recruitment vs. Enrollment

We considered the two stages – recruitment and enrollment – for when a *Michigan Medicine* patient might be excluded from the final *MGI Enrolled* population. To compare the extent to which each stage independently contributed to differences between the two cohorts, we measured the relative participation rate of each stage while controlling for additional impact of the other stage using an approach based on resampling-based simulations.

1. To measure the effect of enrollment, we aim to calculate a subgroup's relative participation rate of being in *MGI Enrolled* given no disparity is introduced in the recruitment stage. To simulate what would have happened in this scenario, we first sample the *MGI Recruited* population from the *Michigan Medicine* population using equal recruitment rates for each subgroup (e.g., male and female). From this resampled *MGI Recruited* population, we then sample the *MGI Enrolled* population according to the observed enrollment rates for each subgroup. Finally, we calculate the relative participation rate of the subgroup of interest using the resampled *MGI Enrolled* population.
2. To measure the effect of recruitment, we aim to calculate a subgroup's relative participation rate of being in *MGI Enrolled* given no disparity is introduced in the enrollment stage. To simulate what would have happened in this scenario, we first sample from the *Michigan Medicine* population according to the observed recruitment rates for each subgroup to create the *MGI Recruited* population, and then sample from this population using equal enrollment rates for each subgroup to create the *MGI Enrolled* population. Finally, we calculate the relative participation rate of the subgroup of interest using the resampled *MGI Enrolled* population.

We repeated the simulations for 1,000 runs. We compared the relative participation rates using a two-tailed z-test with a Bonferroni correction for multiple hypotheses ( $\alpha_2 = 0.001$ ).

## Supplementary Results

### Unknown Socioeconomic Index

86.1%, 90.3%, and 90.1% of the patient addresses from *Michigan Medicine*, *MGI Recruited*, and *MGI Enrolled*, respectively, were successfully geocoded.

### Unknown Race & Ethnicity

In *Michigan Medicine*, unknown race or unknown ethnicity was reported for a large proportion of patients (17.8% and 20.63%, respectively). Here, we consider various assumptions of how patients with unknown race or ethnicity are distributed and report how these assumptions affect our main conclusions.

Race:

- If we assume all patients with unknown race are White, then the proportions of White vs. Black or African American patients are 87.6% vs. 7.3% in *Michigan Medicine*, and 94.8% vs. 5.1% in *MGI Enrolled*.
- If we assume all patients with unknown race are African American, then the proportions of White vs. Black or African American patients are 69.8% vs. 25.1% in *Michigan Medicine* and 89.7% vs. 8.2% in *MGI Enrolled*.
- If we assume the patients with unknown race are equally distributed with respect to known races, then the proportions of White vs. Black or African American patients are 84.9% vs. 8.9% in *Michigan Medicine* and 92.6% vs. 5.3% in *MGI Enrolled*.

Under these distribution assumptions of unknown race, the main conclusions in the results of the main text hold: White patients are overrepresented and Black or African American patients are underrepresented in *MGI Enrolled* compared to *Michigan Medicine*.

Ethnicity:

- If we assume all patients with unknown ethnicity are Hispanic, then the proportions of Hispanic vs. non-Hispanic patients are 22.92% vs. 77.08% in *Michigan Medicine* and 5.27% vs. 94.73% in *MGI Enrolled*.
- If we assume all patients with unknown ethnicity are non-Hispanic, then the proportions of Hispanic vs. non-Hispanic patients are 2.29% vs. 97.71% in *Michigan Medicine* and 1.93% vs. 98.07% in *MGI Enrolled*.
- If we assume the patients with unknown ethnicity are equally distributed with respect to known ethnicities, then the proportions of Hispanic vs. non-Hispanic patients are 2.89% vs. 97.11% in *Michigan Medicine* and 2.00% vs. 98.00% in *MGI Enrolled*.

Under these distribution assumptions of unknown ethnicity, the main conclusions of the results of the main text hold: non-Hispanic patients are overrepresented and Hispanic patients are underrepresented in *MGI Enrolled* compared *Michigan Medicine*.

**Table S1. Study cohort characteristics.**

	<b>Michigan Medicine</b>	<b>MGI Recruited</b>	<b>MGI Enrolled</b>
<b>N</b>	1,242,826	95,206	67,687
<b>Age</b>			
median [IQR]	50.7 [32.3-65.2]	56.2 [42.2-66.9]	56.0 [42.5-66.4]
<b>Sex (%)</b>			
• Female	695,823 (56.0)	51,121 (53.7)	36,328 (53.7)
• Male	546,871 (44.0)	44,082 (46.3)	31,357 (46.3)
• Unknown	132 (0.011)	3 (0.003)	2 (0.003)
<b>Race (%)</b>			
• White	867,791 (69.8)	82,571 (86.7)	60,733 (89.7)
• Black or African American	90,933 (7.3)	6,623 (7.0)	3,473 (5.1)
• Asian	58,365 (4.7)	1,987 (2.1)	1,031 (1.5)
• Other	225,737 (18.2)	4,025 (4.2)	2,450 (3.6)
* Other race includes: Native Hawaiian and other Pacific Islander, American Indian and Alaska Native, Other, Unknown, Patient Refused			
<b>Ethnicity (%)</b>			
• Hispanic	28,450 (2.3)	2,078 (2.2)	1,307 (1.9)
• Non-Hispanic	957,983 (77.1)	89,796 (94.3)	64,117 (94.7)
• Other	256,393 (20.6)	3,332 (3.5)	2,263 (3.3)
* Other ethnicity includes: Unknown, Patient Refused			

**Table S2. (a) Summary of age, sex, race, ethnicity, and (b) neighborhood disadvantage index (overall + stratified by race and ethnicity) across the three cohorts.**

<b>(a) Demographics</b>	<b>MM</b>	<b>MGI-Recr.</b>	<b>MGI-Enr.</b>
<b>N</b>	1,242,826	95,206	67,687
<b>Age, median (IQR)</b>	50.7 (32.3-65.2)	56.2 (42.3-66.9)	56.0 (42.5-66.4)
<b>Sex (%)</b>			
• Male	44.00	46.30	46.33
• Female	55.99	53.70	53.67
• Unknown	0.011	0.003	0.003
<b>Race (%)</b>			
• White	69.82	86.73	89.73
• Black or African American	7.32	6.96	5.13
• Asian	4.70	2.09	1.52
• American Indian and Alaska Native	0.31	0.44	0.46
• Native Hawaiian and other Pacific Islander	0.080	0.074	0.071
• Other, Unknown, Patient Refused	17.77	3.72	3.09
<b>Ethnicity (%)</b>			
• Hispanic	2.29	2.18	1.93
• Non-Hispanic	77.08	94.32	94.73
• Unknown, Patient Refused	20.63	3.50	3.34
<b>(b) Socioeconomic Status: Neighborhood Disadvantage Index</b>			
	<b>MM</b>	<b>MGI-Recr.</b>	<b>MGI-Enr.</b>
<b>Number of patients successfully mapped (%)</b>	1,070,442 (86.1%)	85,995 (90.3%)	60,981 (90.1%)
<b>Overall, median (IQR)</b>	0.074 (0.042-0.122)	0.073 (0.042-0.116)	0.071 (0.042-0.113)
<b>By Race, median (IQR)</b>			
• White	0.067 (0.040-0.107)	0.069 (0.042-0.107)	0.069 (0.042-0.106)
• Black or African American	0.148 (0.085-0.258)	0.144 (0.080-0.247)	0.138 (0.076-0.243)
• Asian, AIAN, NHPI, Other, Unknown, Patient Refused	0.081 (0.046-0.137)	0.075 (0.040-0.123)	0.074 (0.039-0.120)
<b>By Ethnicity, median (IQR)</b>			
• Hispanic	0.093 (0.052-0.151)	0.092 (0.053-0.151)	0.084 (0.050-0.137)
• Non-Hispanic	0.072 (0.041-0.119)	0.074 (0.043-0.117)	0.073 (0.043-0.114)
• Unknown, Patient Refused	0.077 (0.045-0.130)	0.068 (0.040-0.112)	0.067 (0.039-0.107)

Acronyms: IQR, interquartile range; MM, *Michigan Medicine*; MGI-Recr., *MGI Recruited*; MGI-Enr., *MGI Enrolled*.

**Table S3. Comparison and hypothesis testing of age, sex, race, ethnicity, and socioeconomic status (overall + stratified by race and ethnicity) across the three cohorts.**

	MGI-Enr. vs MM		MGI-Recr. vs MM		MGI-Enr. vs MGI-Recr.	
	effect size	p-value	effect size	p-value	effect size	p-value
<b>Age</b>	57.7%	< 0.001	57.0%	< 0.001	49.6%	0.005
<b>Sex</b>						
• Male	2.23 (1.79, 2.85)	< 0.001	2.30 (1.85, 2.75)	< 0.001	0.02 (-0.65, 0.70)	0.92
• Female	-2.23 (-2.85, -1.79)	< 0.001	-2.30 (-2.74, -1.84)	< 0.001	-0.02 (-0.70, 0.65)	0.92
• Unknown	-0.01 (-0.01, -0.00)	0.05	-0.01 (-0.01, -0.00)	0.03	-0.00 (-0.01, 0.01)	0.94
<b>Race</b>						
• White	19.90 (19.57, 20.24)	< 0.001	16.90 (16.59, 17.22)	< 0.001	3.00 (2.57, 3.43)	< 0.001
• Black or African American	-2.19 (-2.42, -1.95)	< 0.001	-0.36 (-0.59, -0.13)	< 0.001	-1.83 (-2.14, -1.51)	< 0.001
• Asian	-3.17 (-3.31, -3.04)	< 0.001	-2.61 (-2.74, -2.47)	< 0.001	-0.56 (-0.74, -0.39)	< 0.001
• American Indian and Alaska Native	0.14 (0.06, 0.21)	< 0.001	0.13 (0.07, 0.18)	< 0.001	0.02 (-0.07, 0.11)	0.58
• Native Hawaiian and other Pacific Islander	-0.01 (-0.04, 0.02)	0.49	-0.01 (-0.03, 0.02)	0.59	-0.00 (-0.04, 0.03)	0.85
• Other, Unknown, Patient Refused	-14.68 (-14.88, -14.48)	< 0.001	-14.06 (-14.24, -13.87)	< 0.001	-0.62 (-0.87, -0.38)	< 0.001
<b>Ethnicity</b>						
• Hispanic	-0.36 (-0.50, -0.21)	< 0.001	-0.11 (-0.24, 0.03)	0.03	-0.25 (-0.44, -0.06)	< 0.001
• Non-Hispanic	17.64 (17.39, 17.90)	< 0.001	17.24 (17.01, 17.46)	< 0.001	0.41 (0.10, 0.71)	< 0.001
• Unknown, Patient Refused	-17.29 (-17.50, -17.08)	< 0.001	-17.13 (-17.32, -16.94)	< 0.001	-0.16 (-0.40, 0.09)	0.09
<b>Neighborhood socioeconomic disadvantage index</b>						
• Overall	48.5%	< 0.001	49.3%	< 0.001	49.2%	< 0.001
• Race: White	50.8%	< 0.001	51.0%	< 0.001	49.7%	0.11
• Race: Black or African American	47.8%	< 0.001	48.8%	0.00199	49.0%	0.13
• Race: Asian, AIAN, NHPI, Other, Unknown, Patient Refused	45.3%	< 0.001	46.2%	< 0.001	49.2%	0.17
• Ethnicity: Hispanic	47.2%	0.00495	49.9%	0.88	47.3%	0.03
• Ethnicity: Non-Hispanic	50.1%	0.47765	50.8%	< 0.001	49.3%	0.00004
• Ethnicity: Unknown, Patient Refused	43.6%	< 0.001	44.5%	< 0.001	49.1%	0.00089

Acronyms: MM, *Michigan Medicine*; MGI-Recr., *MGI Recruited*; MGI-Enr., *MGI Enrolled*.

The effect size for Mann-Whitney U test is defined as the proportion of pairs between the two cohorts that support the alternative hypothesis (where an effect size closer to 50% is likely not statistically significant). Difference in proportions is reported for sex, race, and ethnicity with confidence intervals adjusted for the Bonferroni correction (99.7%).

**Table S4. Comparisons of consent/decline rates across subgroups. (Continuous variables are split by median value of *Michigan Medicine*)**

	Consent rate	Decline rate
<b>Overall</b>	71.1%	28.9%
<b>Age</b>		
• < 50.7	71.3%	28.7%
• ≥ 50.7	70.9%	29.0%
<b>Sex</b>		
• Male	71.1%	28.9%
• Female	71.1%	28.9%
• Unknown	66.7%	33.3%
<b>Race</b>		
• White	73.6%	26.4%
• Black or African American	52.4%	47.6%
• Asian	51.9%	48.1%
• American Indian and Alaska Native	74.1%	25.9%
• Native Hawaiian and other Pacific Islander	68.6%	31.4%
• Other, Unknown, Patient Refused	59.2%	40.8%
<b>Ethnicity</b>		
• Hispanic	62.9%	37.1%
• Non-Hispanic	71.4%	28.6%
• Unknown, Patient Refused	67.9%	32.1%
<b>Neighborhood socioeconomic disadvantage index</b>		
• < 0.074	72.5%	27.5%
• ≥ 0.074	69.3%	30.7%

**Table S5. Comparison of overall inclusion rate, recruitment rate, and consent rate across demographic subgroups, listed as the raw fraction and normalized with respect to rates of the total cohort.**

	Overall inclusion rate	1) Recruitment rate	2) Consent rate	Normalized overall inclusion rate	1) Normalized recruitment rate	2) Normalized consent rate
<b>Total</b>	5.4%	7.7%	71.1%	1	1	1
<b>Sex</b>						
• Male	5.7%	8.1%	71.1%	1.05	1.05	1.00
• Female	5.2%	7.3%	71.1%	0.96	0.96	1.00
• Unknown	1.5%	2.3%	66.7%	0.28	0.30	0.94
<b>Race</b>						
• White	7.0%	9.5%	73.6%	1.29	1.24	1.03
• Black or African American	3.8%	7.3%	52.4%	0.70	0.95	0.74
• Asian	1.8%	3.4%	51.9%	0.32	0.44	0.73
• American Indian and Alaska Native	8.0%	10.7%	74.1%	1.46	1.40	1.04
• Native Hawaiian and other Pacific Islander	4.9%	7.2%	68.6%	0.90	0.94	0.96
• Other, Unknown, Patient Refused	0.9%	1.6%	59.2%	0.17	0.21	0.83
<b>Ethnicity</b>						
• Hispanic	4.6%	7.3%	62.9%	0.84	0.95	0.88
• Non-Hispanic	6.7%	9.4%	71.4%	1.23	1.22	1.00
• Unknown, Patient Refused	0.9%	1.3%	67.9%	0.16	0.17	0.96
<b>Socioeconomic Status Index</b> (stratified by median value of <i>Michigan Medicine</i> )						
• Disadvantage index < 0.074	5.9%	8.1%	72.5%	1.08	1.06	1.02
• Disadvantage index ≥ 0.074	5.5%	7.9%	69.3%	1.01	1.04	0.97
• Unknown	3.9%	5.3%	72.8%	0.71	0.70	1.02

A normalized rate of 1 means the subgroup is included (or recruited/consented) at the same rate as the entire cohort; a normalized rate of <1 or >1 means the subgroup is included (or recruited/consented) at a higher rate than the entire cohort.

**Table S6. Comparison of relative risk of MGI participation by demographic subgroups, and a breakdown of disparity from the recruitment stage and the consent stage.**

	MM→MGI-Enr. Overall disparity of MGI participation	(1) MM→MGI-Recr. Disparity due to recruitment	(2) MGI-Recr.→MGI-Enr. Disparity due to consent
<b>Sex</b>			
• Male	1.1	1.10* (1.07 - 1.13)	1.00 (0.98 - 1.02)
• Female	0.91	0.91* (0.89 - 0.94)	1.00 (0.98 - 1.02)
• Unknown	0.28	0.30* (0.00 - 1.32)	0.94 (0.07 - 2.36)
<b>Race</b>			
• White	3.77	2.82* (2.72 - 2.93)	1.34 (1.30 - 1.38)
• Black or African American	0.69	0.95 (0.90 - 1.00)	0.72* (0.69 - 0.77)
• Asian	0.31	0.43* (0.40 - 0.47)	0.73 (0.68 - 0.77)
• American Indian and Alaska Native	1.46	1.40* (1.13 - 1.68)	1.04 (0.82 - 1.23)
• Native Hawaiian and other Pacific Islander	0.9	0.94* (0.52 - 1.41)	0.96 (0.58 - 1.41)
• Other, Unknown, Patient Refused	0.15	0.18* (0.17 - 0.19)	0.83 (0.80 - 0.85)
<b>Ethnicity</b>			
• Hispanic	0.84	0.95 (0.89 - 1.03)	0.88* (0.80 - 0.96)
• Non-Hispanic	5.34	4.94* (4.67 - 5.23)	1.08 (1.05 - 1.12)
• Unknown, Patient Refused	0.13	0.14* (0.13 - 0.15)	0.95 (0.93 - 0.98)
<b>Socioeconomic Status Index</b> (stratified by median value of <i>Michigan Medicine</i> )			
• Disadvantage index < 0.074	1.15	1.11* (1.09 - 1.14)	1.04 (1.01 - 1.06)
• Disadvantage index ≥ 0.074	1.02	1.07* (1.05 - 1.10)	0.96 (0.93 - 0.98)
• Unknown	0.68	0.67* (0.64 - 0.70)	1.03 (0.99 - 1.06)

Acronyms: RR, relative rate of participation/recruitment/consent; MM, *Michigan Medicine*; MGI-Recr., *MGI Recruited*; MGI-Enr., *MGI Enrolled*. RR<1 means being that subgroup is less likely to participate in MGI; RR>1 means being that subgroup is more likely to participate in MGI. Comparing the two stages (1) and (2), an RR further away from 1 (smaller if RR<1, larger if RR>1) has a relatively greater effect on the overall difference between *MM* and *MGI Enrolled* (\*starred results indicate that the difference from that stage is significantly larger than the other stage). Confidence intervals are adjusted for the Bonferroni correction (99.9%).



**Table S7. Relative risk comparisons for recruitment and consent stages for socioeconomic indices, controlling for race and ethnicity.**

	MM→MGI-Enr. Overall disparity	(1) MM→MGI-Recr. Disparity due to recruitment	(2) MGI-Recr.→MGI-Enr. Disparity due to consent
<b>Race: White</b>			
• Disadvantage index < 0.067	0.941	0.927* (0.901 - 0.951)	1.016 (0.993 - 1.040)
• Disadvantage index ≥ 0.067	1.063	1.079* (1.052 - 1.106)	0.985 (0.961 - 1.006)
<b>Race: Black or African American</b>			
• Disadvantage index < 0.149	1.160	1.084* (0.976 - 1.199)	1.070 (0.962 - 1.183)
• Disadvantage index ≥ 0.149	0.862	0.923* (0.828 - 1.017)	0.935 (0.841 - 1.035)
<b>Race: Asian</b>			
• Disadvantage index < 0.067	1.481	1.427* (1.190 - 1.713)	1.038 (0.878 - 1.265)
• Disadvantage index ≥ 0.067	1.161	1.209* (1.025 - 1.467)	0.960 (0.803 - 1.158)
<b>Race: AIAN, NHPI, Other, Unknown, Patient Refused</b>			
• Disadvantage index < 0.084	1.239	1.139* (1.040 - 1.267)	1.088 (0.990 - 1.191)
• Disadvantage index ≥ 0.084	0.807	0.878* (0.796 - 0.971)	0.919 (0.839 - 1.014)
<b>Ethnicity: Hispanic</b>			
• Disadvantage index < 0.093	1.251	1.033 (0.840 - 1.263)	1.211* (1.003 - 1.451)
• Disadvantage index ≥ 0.093	0.800	0.968 (0.796 - 1.184)	0.826* (0.683 - 0.994)
<b>Ethnicity: Non-Hispanic</b>			
• Disadvantage index < 0.072	0.972	0.936* (0.910 - 0.958)	1.039 (1.009 - 1.070)
• Disadvantage index ≥ 0.072	1.029	1.069* (1.040 - 1.095)	0.963 (0.934 - 0.992)
<b>Ethnicity: Unknown, Patient Refused</b>			
• Disadvantage index < 0.077	1.428	1.342* (1.282 - 1.395)	1.064 (1.025 - 1.114)
• Disadvantage index ≥ 0.077	0.700	0.745* (0.715 - 0.778)	0.940 (0.906 - 0.982)

Acronyms: RR, relative risk; MM, *Michigan Medicine*; MGI-Recr., *MGI Recruited*; MGI-Enr., *MGI Enrolled*.

RR<1 means being that subgroup is less likely to participate in MGI; RR>1 means being that subgroup is more likely to participate in MGI. Comparing the two stages (1) and (2), an RR further away from 1 (smaller if RR<1, larger if RR>1) has a relatively greater effect on the overall difference between *MM* and *MGI Enrolled* (\*starred results indicate that the difference from that stage is significantly larger than the other stage). Confidence intervals are adjusted for the Bonferroni correction (99.6%).

**Table S8. Comparison of relative risk of MGI participation by racial subgroups (relative to White), and a breakdown of disparity from the recruitment stage and the consent stage.**

	MM→MGI-Enr. Overall disparity vs White	(1) MM→MGI-Recr. Disparity due to recruitment vs White	(2) MGI-Recr.→MGI-Enr. Disparity due to consent vs White
<b>Race</b>			
• White	1	1	1
• Black or African American	0.55	0.77 (0.73 - 0.81)	0.71* (0.67 - 0.75)
• Asian	0.25	0.36* (0.32 - 0.40)	0.71 (0.66 - 0.75)
• American Indian and Alaska Native	1.14	1.13* (0.89 - 1.38)	1.01 (0.84 - 1.21)
• Native Hawaiian and other Pacific Islander	0.70	0.75* (0.42 - 1.16)	0.93 (0.59 - 1.32)
• Other, Unknown, Patient Refused	0.14	0.17* (0.15 - 0.18)	0.80 (0.79 - 0.84)

Acronyms: RR, relative rate of participation/recruitment/consent; MM, *Michigan Medicine*; MGI-Recr., *MGI Recruited*; MGI-Enr., *MGI Enrolled*.

RR<1 means being that subgroup is less likely to participate in MGI; RR>1 means being that subgroup is more likely to participate in MGI. Comparing the two stages (1) and (2), an RR further away from 1 (smaller if RR<1, larger if RR>1) has a relatively greater effect on the overall difference between *MM* and *MGI Enrolled* (\*starred results indicate that the difference from that stage is significantly larger than the other stage). Confidence intervals are adjusted for the Bonferroni correction (99.9%).