Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Ahdoot M, Wilbur AR, Reese SE, et al. MRI-targeted, systematic, and combined biopsy for prostate cancer diagnosis. N Engl J Med 2020;382:917-28. DOI: 10.1056/NEJMoa1910038

APPENDIX

TABLE OF CONTENTS

Authors' Contributions/ Trio Study Members

Drs. Ahdoot and Pinto had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design*:* Ahdoot, Choyke, Parnes, Pinto, Turkbey, Linehan, Shih, Wood.

Acquisition, analysis, or interpretation of data*:* Ahdoot, Bloom, Choyke, Gomella, Gurram, Lebastchi, Linehan, Mehralivand, Merino, Parnes, Pinsky, Pinto, Reese, Shih, Siddiqui, Turkbey, Wilbur, Wood

Drafting of the manuscript*:* Ahdoot, Bloom, Choyke, Gomella, Gurram, Lebastchi, Linehan, Mehralivand, Merino, Parnes, Pinsky, Pinto, Reese, Shih, Siddiqui, Turkbey, Wilbur, Wood

Critical revision of the manuscript for important intellectual content*:* Ahdoot, Bloom, Choyke, Gomella, Gurram, Lebastchi, Linehan, Mehralivand, Merino, Parnes, Pinsky, Pinto, Reese, Shih, Siddiqui, Turkbey, Wilbur, Wood

Statistical analysis*:* Pinsky, Reese, Shih

Obtained funding*:* Choyke, Linehan, Pinto, Wood.

Administrative, technical, or material support*:* Choyke, Linehan, Merino, Parnes, Pinsky, Pinto, Siddiqui, Turkbey, Wood.

Study supervision*:* Choyke, Linehan, Parnes, Pinto, Shih, Turkbey, Wood.

Other study contributors*:* The authors would like to recognize the other members of the Trio Study team for their consistent and outstanding dedication to clinical care and scientific research. Thank you to Paul Wakim, Victoria Anderson, Charisse Garcia, Julie Peretti, Mabel Cruz, Charlotte Payne, Nana Yaqub-Ogun, Michele Reed, Toneisia Gross, Lerkia Parks, Luis Nunez, Bryant Villavicencio, Dagane Daar, Tieu Hoa, Yolanda McKinney, Juanita Weaver, and Sheng Xu. A special thank you to the patients and their families that participated in our clinical trial.

No commercial entities were involved in the authorship of this manuscript.

Description of Industry Collaboration

The electromagnetic tracking device and software used to perform MRI-targeted biopsies described in this manuscript was developed through a collaborative research and development agreement (CRADA) between the National Institutes of Health (NIH) and Philips Medical Systems, Inc. Philips provided the NIH with materials and technical expertise to help facilitate the project. Koninklijke Philips has licensed patent rights from the National Institutes of Health.

Some of the authors have intellectual property in fields related to the topic of this manuscript for which they receive royalties. Financial disclosures provided by each author are available at NEJM.org.

Detailed Description of Statistical Methods

All patients enrolled in the study underwent MRI-targeted and systematic prostate biopsies. The results of these prostate biopsies were reported as Gleason scores and coded as the highest Grade Group detected by each biopsy method. The primary outcomes were the rates of cancer detection by each biopsy method reported by Grade Groups. Biopsies resulted in no cancer detection or one of five possible prostate cancer Grade Groups.[1] Grade Groups range from a score of 1, which signifies the lowest risk disease, to 5, which signifies the highest risk prostate cancer.[2] Given the five possible primary outcomes (GG 1-5) a Bonferroni correction was used to correct for multiplicity and two-sided p-values were calculated using McNemar's test, with a p-value of <0.01 (instead of $p<0.05$) being considered statically significant. Confidence Intervals for cancer detection rates by each biopsy method and differences in cancer detection rates between the two biopsy methods were calculated using the adjusted Wald interval for each Grade Group and were widened (from 95%) to 99% confidence intervals to account for the 5 possible primary outcomes.[3,4] Differences in proportions of Grade Group cancer detection by each biopsy method are shown in a graphical presentation in Figure 2 and a tabular presentation in Table S2.

Secondary outcomes were Grade Group ≥3 (and Grade Group ≥2) cancer detection rates, cancer detection rates stratified by prior biopsy status, and reclassification rates from biopsy to whole-mount histopathology at prostatectomy. For the secondary analysis, p-values calculated using McNemar's test were reported for the comparison between the two biopsy methods for Grade Group ≥3 detection rates (Table S2), Grade Group ≥2 detection rates (Table S2), and all six reclassification rates (Table S5). Given the 8 possible secondary outcomes, a Bonferroni correction was used to correct for multiple comparisons of these 8 secondary outcomes and two-

sided p-values of <0.006 were considered statically significant. Similarly, the adjusted Wald intervals were widened to 99.4% to control for multiplicity.

For the secondary outcome of cancer detection by prior biopsy status (Table S3), the rates of cancer detection by systematic or targeted biopsy were compared between the prior biopsy and biopsy naïve sub-cohorts to estimate if the rates detection for each biopsy method changes based on prior biopsy status. Differences in cancer detection rates between biopsy-naïve and prior biopsy cohorts by any particular biopsy method were defined as the cancer detection rate in the biopsy-naïve cohort minus the rate found in the prior biopsy cohort. 95% confidence intervals for the cancer detection rates with respect to prior biopsy status and difference in cancer detection rates between the two sub-cohorts were calculated using the adjusted Wald intervals (Table S3). P-values were not reported.

Quality Control

All data were input by a dedicated data manager. Before data analysis, a random audit with a 10% sampling rate of the dataset was performed by a urologist (M.A.) to ensure data accuracy. An error rate of greater than 0.5% upon an audit of the data was deemed acceptable and was not reached. This data audit was done in addition to the standard institutional review board data monitoring protocols for our institution.

Table S1

PI-RADS
PI-RADS 1 = Very low (clinically significant cancer highly unlikely)
PI-RADS 2 = Low (clinically significant cancer unlikely)
PI-RADS 3 = Intermediate (clinically significant cancer equivocal)
PI-RADS 4 = High (clinically significant cancer likely)
PI-RADS 5 = Very high (clinically significant cancer highly likely)

Table S1: Summary of PI-RADS Scoring System

PI-RADS = Prostate Imaging Reporting & Data System
Image source: Prostate Cancer - PI-RADS v2. The Radiology Assistant : Prostate Cancer - PI-RADS v2. https://radiologyassistant.nl/abdomen/prostate-cancer-pi-rads-v2. Accessed November 20, 2019.

Table S2

	Grade Group	Systematic (n)	Targeted (n)	Combined (n)	Difference (CI) in Cancer Detection Rates (Targeted - Systematic)	p-value (Targeted vs Systematic)
Gleason Grade Group Detection by Biopsy Method	No Cancer	47.5 (999)	48.5 (1019)	37.6 (791)	1 (-1.6, 3.5)	-
	1	21.6 (454)	13.7 (289)	18.7 (394)	-7.8 (-10.3, -5.3)	<0.001
	2	17.1 (359)	17.6 (370)	21.5 (452)	0.5 (-1.8, 2.9)	0.61
	3	3.5 (73)	5.1 (108)	5.9 (124)	1.7 (0.2, 3.1)	0.004
	4	6.5 (137)	10.2 (215)	10.8 (228)	3.7 (2.2, 5.2)	<0.001
	5	3.9 (81)	4.9 (102)	5.4 (114)	1 (0.2, 1.8)	0.003
	Any Cancer (GG 1 -5)	52.5 (1104)	51.5 (1084)	62.4 (1312)	-1 (-2.9, 1.0)	
Additional Clinically Significant Cancer Diagnosis by Biopsy Method	GG ≥ 2	5.8 (123)	12.7 (268)	-	6.9 (4.3, 9.4)	< 0.001
	GG ≥ 3	1.9 (41)	8.3 (175)	-	6.4 (4.5, 8.3)	<0.001

Table S2: Cancer Detection Rates by Biopsy Method

Reported are the percentage (and number) of cancers detected by grade group and additional clinically significant cancer diagnosis by biopsy method. Differences in cancer detection rates were defined as the detection rate for targeted biopsy minus systematic biopsy and reported with confidence intervals. Given the five possible grade groups (GG 1-5) for the primary outcome, a Bonferroni correction was used leading to p-value <0.01 being considered statistically significant and confidence intervals to be set to 99% for Gleason Grade Group Detection by Biopsy Method. Additional clinically significant cancer diagnosis by biopsy method (shown in yellow) was defined as any newly diagnosed GG≥2 or GG≥3 cancer diagnosed by the addition of the second listed biopsy method. Given eight secondary comparisons were performed a Bonferroni correction was used

and a two-sided p-value <0.006 was considered statistically significant and confidence interval was set to 99.4% for the secondary endpoint of additional clinically significant cancer diagnosis by biopsy method.

CI = Confidence Interval.

Table S3

Differences in Cancer Detection Rates by Prior Biopsy Status

		Systematic vs Systematic	Targeted vs Targeted
		Difference in Cancer Detection Rate (Biopsy Naïve - Prior Biopsy)	Difference in Cancer Detection Rate (Biopsy Naïve - Prior Biopsy)
Additional Clinically Significant Cancer Diagnosis	GG ≥ 2	-0.2% (-2.5, 2.5)	-5.9% (-8.9, -2.7)
	GG ≥ 3	-0.4% (-1.7, 1.1)	-0.7% (-3.4, 2.4)

Table S3: Cancer Detection by Prior Biopsy Status.

Percentage (and 95% confidence interval) for differences in cancer detection rates for systematic or targeted biopsy by prior biopsy status are reported. Differences in cancer detection rates between biopsy-naïve and prior biopsy cohorts by any particular biopsy method were defined as the cancer detection rate in the biopsy-naïve cohort minus the rate found in the prior biopsy cohort. The table demonstrates similar additional GG≥3 cancer detection by systematic or targeted biopsy between biopsy-naïve and prior biopsy patients. No adjustment for multiple comparisons was made and p-values were not reported. Additional GG≥2 or GG≥3 cancer detection was not reported for combined biopsy as cancer detection by combined biopsy was defined as the highest GG detected by either constituent biopsy method.

Table S4

Crosstabulations of Highest Grade Group Detected by Biopsy Method

		MRI-Targeted Biopsy							Systematic Biopsy							Combined Biopsy					
	GG by Biopsy	0	1	2	3	4	5	Sum	0	1	2	3	4	5	Sum	1	2	3	4	5	Sum
Grade Group on Prostatectomy Whole-Mount	GG 1	7 (1.7)	10 (2.5)	8 (2)	1 (0.2)	1 (0.2)	0 (0)	27 (6.7)	2 (0.5)	16 (4)	8 (2)	1 (0.2)	0 (0)	0 (0)	27 (6.7)	12 (3)	12 (3)	2 (0.5)	1 (0.2)	0 (0)	27 (6.7)
	GG 2	24 (5.9)	38 (9.4)	123 (30.4)	15 (3.7)	18 (4.5)	1 (0.2)	219 (54.2)	20 (5)	62 (15.3)	111 (27.5)	13 (3.2)	12 (3)	1 (0.2)	219 (54.2)	24 (5.9)	148 (36.6)	24 (5.9)	22 (5.4)	1 (0.2)	219 (54.2)
	GG 3	1 (0.2)	1 (0.2)	12 (3)	16 (4)	15 (3.7)	0 (0)	45 (21.5)	5 (1.2)	6 (1.5)	9 (2.2)	15 (3.7)	10 (2.5)	0 (0)	45 (11.1)	0 (0)	5 (1.2)	21 (5.2)	19 (4.7)	0 (0)	45 (11.1)
	GG 4	3 (0.7)	7 (1.7)	11 (2.7)	12 (3)	50 (12.4)	4 (1)	87 (6.4)	12 (3)	10 (2.5)	16 (4)	9 (2.2)	39 (9.7)	1 (0.2)	87 (21.5)	3 (0.7)	6 (1.5)	13 (3.2)	60 (14.9)	5 (1.2)	87 (21.5)
	GG 5	0 (0)	0 (0)	0 (0)	0 (0)	9 (2.2)	17 (4.2)	26 (6.4)	5 (1.2)	2 (0.5)	3 (0.7)	0 (0)	7 (1.7)	9 (2.2)	26 (6.4)	0 (0)	0 (0)	0 (0)	7 (1.7)	19 (4.7)	26 (6.4)
Sum		35 (8.7)	56 (13.9)	154 (38.1)	44 (10.9)	93 (23)	22 (5.4)	404 (100)	44 (10.9)	96 (23.8)	147 (36.4)	38 (9.4)	68 (16.8)	11 (2.7)	404 (100)	39 (9.7)	171 (42.3)	60 (14.9)	109 (27)	25 (6.2)	404 (100)
Number (%) of Upgrades		35 (8.7)	46 (11.4)	23 (5.7)	12 (3)	9 (2.2)	0 (0)	125 (30.9)	44 (10.9)	80 (19.8)	28 (6.9)	9 (2.2)	7 (1.7)	0 (0)	168 (41.6)	27 (6.7)	11 (2.7)	13 (3.2)	7 (1.7)	0 (0)	58 (14.4)

Key:	
	Upgrading on Whole-Mount Histology
	No Change on Whole-Mount Histology
	Downgrading on Whole-Mount Histology

Table S4: Crosstabulations of Highest Grade Group Detected by Biopsy Against Final Histopathology

Number (and percentage) of men diagnosed with Grade Group 1-5 or no cancer by systematic, MRI-targeted or combined biopsy are shown along the vertical axis. Grade Group diagnosed by whole-mount histopathology is shown along the longitudinal axis. The areas shaded in blue demonstrate men who were downgraded to a lower grade group on whole-mount histopathology. Areas shaded in red demonstrate men who upgraded to higher Grade Group cancer on histopathology relative to the listed biopsy method. By following any particular cell upwards and to the left the reader can find the Grade Group by the biopsy and the final histopathology detected on whole-mount.

Table S5

Rates of Reclassification on Whole Mount Histology

	Systematic	Targeted	Targeted vs Systematic	P-value
Any Upgrading	41.60%	30.90%	-10.6% (-19.5, -1.7)	0.002
Upgrading to ≥ 2	30.20%	18.30%	-11.9% (-19.8, -3.9)	<0.001
Upgrading to ≥ 3	16.80%	8.70%	-8.1% (-13.9, -2.3)	<0.001
Any Downgrading	11.40%	15.60%	4.2% (-1.2, 9.6)	0.044
Downgrading to ≤ 2	6.70%	8.90%	2.2% (-1.9, 6.4)	0.188
Downgrading to ≤ 1	2.20%	2.50%	0.2% (-2.1, 2.6)	1

Table S5: Rates of Reclassification on Whole Mount Histology

Rates of reclassification by downgraded and upgraded on prostatectomy whole-mount histopathology by biopsy method (n=404). Shown are six of the eight secondary comparisons in which with a Bonferroni correction was used. With the use of this correction, a two-sided p-value <0.006 was considered statistically significant and confidence intervals were set to 99.4% and reported in the table above. Any upgrading, Grade Group ≥2, and Grade Group ≥3 upgrading were less common with targeted biopsy and met statistically significance (p<0.006). Downgrading rates by targeted biopsy were higher but not meet statistically significance.

Guide to Interpreting Cross Tabulations

In this manuscript cross-tabulations such as Tables 2 and S4 allow for a dense presentation of patient MRI-targeted and systematic biopsy outcomes. The reader can determine the likelihood of a biopsy outcome by selecting the Grade Group detected by systematic and targeted biopsy to see where these columns and rows intersect. The point of intersection defines the number (and percentage) of patients found to have the prostate cancer with those particular biopsy findings. For example, to determine the number of patients found to have no cancer on systematic biopsy but found to have Grade Group≥3 cancer on MRI-Targeted biopsy, the reader would follow the column labeled "No Cancer" under systematic biopsy down and add the patients found to have Grade Group 3-5 disease on Targeted biopsy(22+29+8=59). Alternatively, by following any particular upwards and to the left the reader can find the Grade Group detected for the patients by systematic and MRI-targeted biopsy, respectively. The sum of these cells will represent the 59 patients out of the total of 2103 patient found to have GG≥3 cancer on targeted biopsy but found to have no cancer on systematic biopsy.

Crosstabulation of Highest Gleason Grade Group Detected by Biopsy Method								
		No Cancer	1	2	3	4	5	Sum
Targeted Biopsy	No Cancer	791 (37.6)	163 (7.8)	56 (2.7)	5 (0.2)	3 (0.1)	1 (0.05)	1019 (48.5)
	1	74 (3.5)	157 (7.5)	50 (2.4)	6 (0.3)	2 (0.1)	0 (0)	289 (13.7)
	2	75 (3.6)	93 (4.4)	178 (8.5)	14 (0.7)	10 (0.5)	0 (0)	370 (17.6)
	3	22 (1)	19 (0.9)	36 (1.7)	22 (1.0)	9 (0.4)	0 (0)	108 (5.1)
	4	29 (1.4)	19 (0.9)	33 (1.6)	25 (1.2)	98 (4.7)	11 (0.5)	215 (10.2)
	5	8 (0.4)	3 (0.1)	6 (0.3)	1 (0.05)	15 (0.7)	69 (3.3)	102 (4.9)
	Sum	999 (47.5)	454 (21.6)	359 (17.1)	73 (3.5)	137 (6.5)	81 (3.9)	2103 (100)

The areas shaded in grey demonstrate men in whom systematic and targeted biopsy detected the same Grade Group. Areas shaded in blue demonstrate men who were found to have higher Grade

Group cancer on targeted biopsy and the areas shaded in green demonstrate men who were found to have higher Grade Group cancer on systematic biopsy. If a reader wished to know the total number of upgrading events associated with the use of Targeted biopsy then a sum of all the cells shaded blue would provide that information. Similarly, added all the cells in green would provide the total number of upgrading events by systematic biopsy. These sums are reported in Figure 2 under the heading “Any Upgrading by Addition of Biopsy Method”.

REFERENCES

1. Epstein JI, Zelefsky MJ, Sjoberg DD, et al. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *Eur Urol.* 2016;69(3):428-435.
2. Epstein JI. Prostate Cancer Grade Groups Correlate with Prostate-specific Cancer Mortality: SEER Data for Contemporary Graded Specimens. *Eur Urol.* 2017;71(5):764-765.
3. Agresti A, Caffo B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *Am Stat.* 2000;54(4):280-288.
4. Agresti A, Min Y. Simple improved confidence intervals for comparing matched proportions. *Stat Med.* 2005;24(5):729-740.