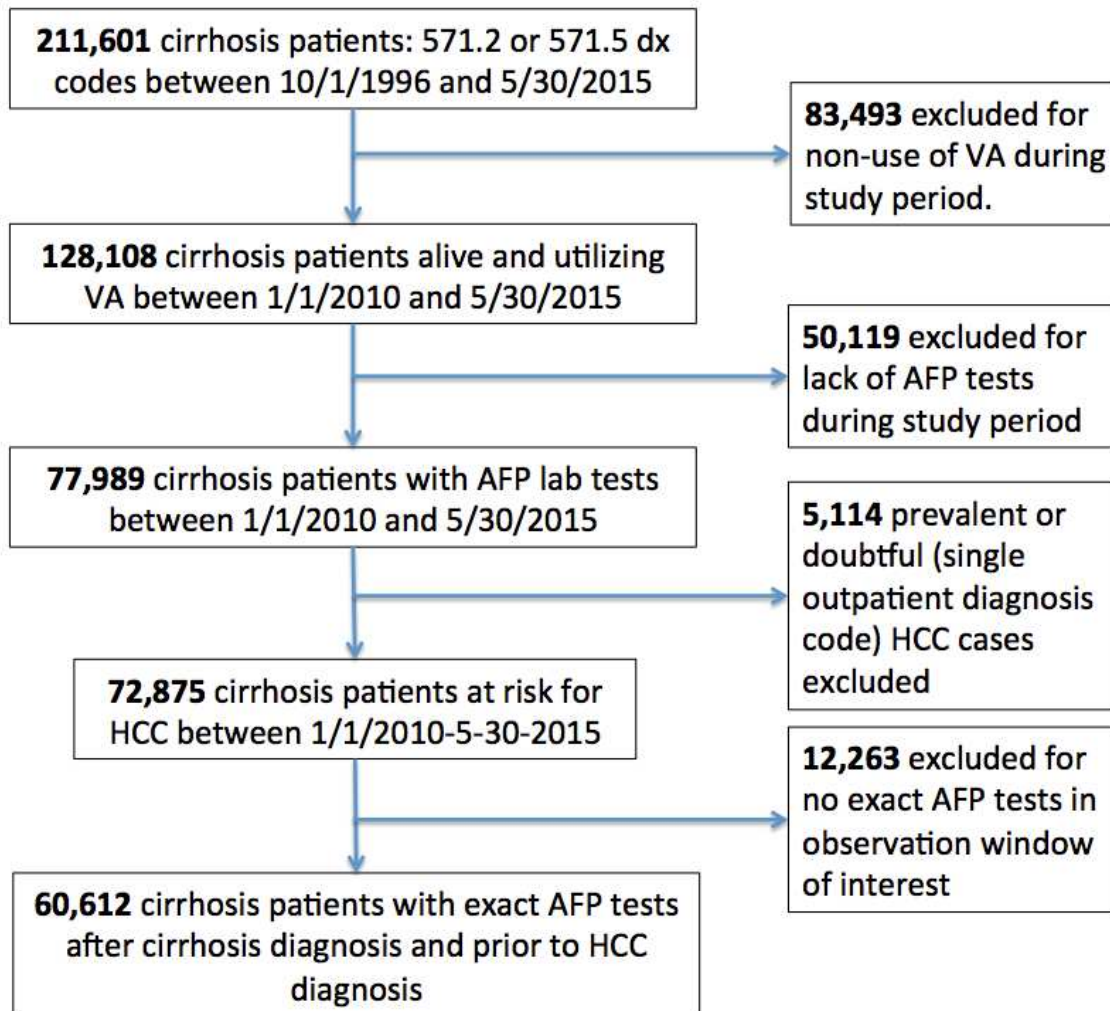1    **Validation of an HCC Early detection Screening (HES) algorithm in a multi-etiology Veteran's**

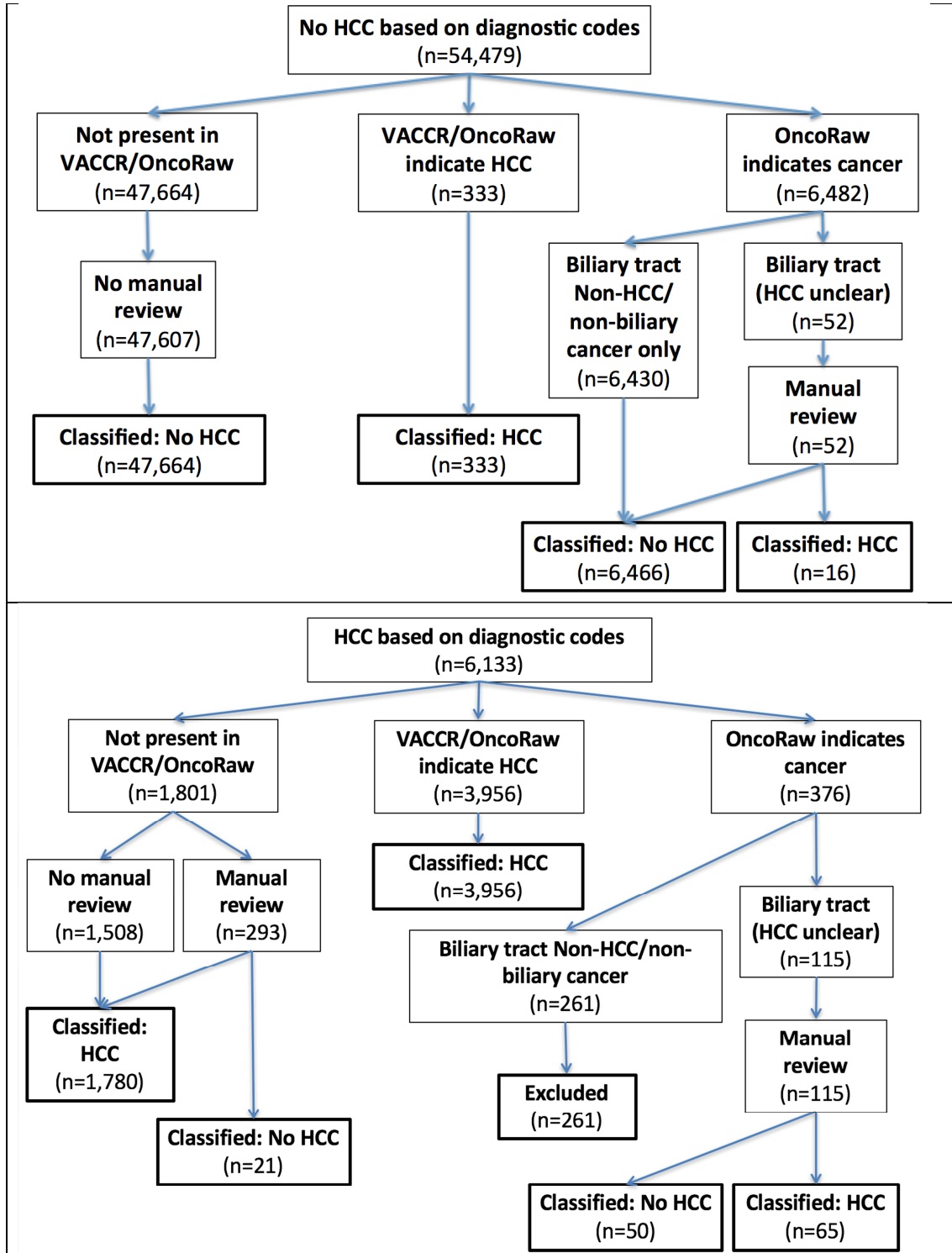2                          **Affairs cirrhosis cohort**

3                          **Additional supplementary materials**

4  **Supplementary Figure 1:** Derivation of the study cohort in the Department of Veteran's Affairs (VA)
5  Corporate Database Warehouse.



6
7
8
9
10

11  **Supplementary Figure 2:** HCC classification procedure in the Department of Veteran's Affairs (VA)
12  cirrhosis cohort.



13  * VACCR: VA Central Cancer Registry; OncoRaw: VA Corporate Database Warehouse oncology raw data
14  files.

15      **Supplementary Table 1:** Demographic table

|  | Controls | HCC cases |
|---|---|---|
| **N** | 33,627 | 4,804 |
| **Age at baseline AFP (years)** | 60.67 (56.18-64.44) | 60.83 (57.22-64.49) |
| **Sex** |  |  |
| Male | 32,575 (96.87%) | 4,743 (98.73%) |
| Female | 1,052 (3.13%) | 61 (1.27%) |
| **Race** |  |  |
| Non-Hispanic White | 21,215 (63.09%) | 2,955 (61.51%) |
| Non-Hispanic Black | 6,705 (19.94%) | 934 (19.44%) |
| Hispanic | 3,210 (9.55%) | 578 (12.03%) |
| Other | 2,497 (7.43%) | 337 (7.01%) |
| **Number of AFP tests** |  |  |
| 1 | 13,528 (40.23%) | 1,801 (37.49%) |
| 2 | 7,265 (21.60%) | 932 (19.40%) |
| 3-4 | 7,677 (22.83%) | 1,173 (24.42%) |
| >4 | 5,157 (15.34%) | 898 (18.69%) |
| **Months between AFP tests** | 6.21 (3.71-11.40) | 5.29 (2.76-8.97) |
| **Baseline AFP (ng/ml)** | 4.40 (2.80-7.97) | 11.01 (5.20-37.23) |
| **Baseline ALT (ng/ml)** | 40 (25-70) | 56 (34-90) |
| **Baseline PLT (1000's)** | 131 (89-183) | 99 (70-138) |
| **Etiology of cirrhosis** |  |  |
| Alcoholic liver disease | 24,137 (71.78%) | 3,737 (77.79%) |
| Hepatitis C virus | 16,264 (48.37%) | 3,392 (70.61%) |
| Hepatitis B virus | 2,962 (8.81%) | 437 (9.10%) |
| Nonalcoholic fatty liver disease | 5,636 (16.76%) | 584 (12.16%) |

16      * Continuous variables: median and interquartile range (25th and 75th percentiles)

17      * Categorical variables: number and percentage

**Definitions of time-dependent patient-level sensitivity and screening-level false positive rate**

**Measure/Biomarker used in screening**: the predicted probability of HCC in the next six months from current AFP test.

**Positive tests**: if the predicted probability of HCC in the next six months exceeds a threshold $c$.

**Sensitivity within T months corresponding to 10% screening level false positive rate**: the proportion of HCC cases with ≥1 positive tests within T months prior to their HCC diagnosis, among all the HCC cases with at least one AFP test within T months of their HCC diagnosis. The threshold $c$ is selected to ensure a 10% screening level false positive rate (T months).

**Screening level false positive test (T months)**: the proportion of positive tests among all tests done in either controls with no HCC or in cases when testing occurs more than 6 months prior to HCC diagnosis.

**Sensitivity between T1 and T2 months corresponding to 10% screening level false positive rate**: the proportion of HCC cases with ≥1 positive tests within T1-T2 months prior to their HCC diagnosis among all the HCC cases with at least one AFP test during this period. The threshold $c$ is selected to ensure a 10% screening level false positive rate (T2 months).
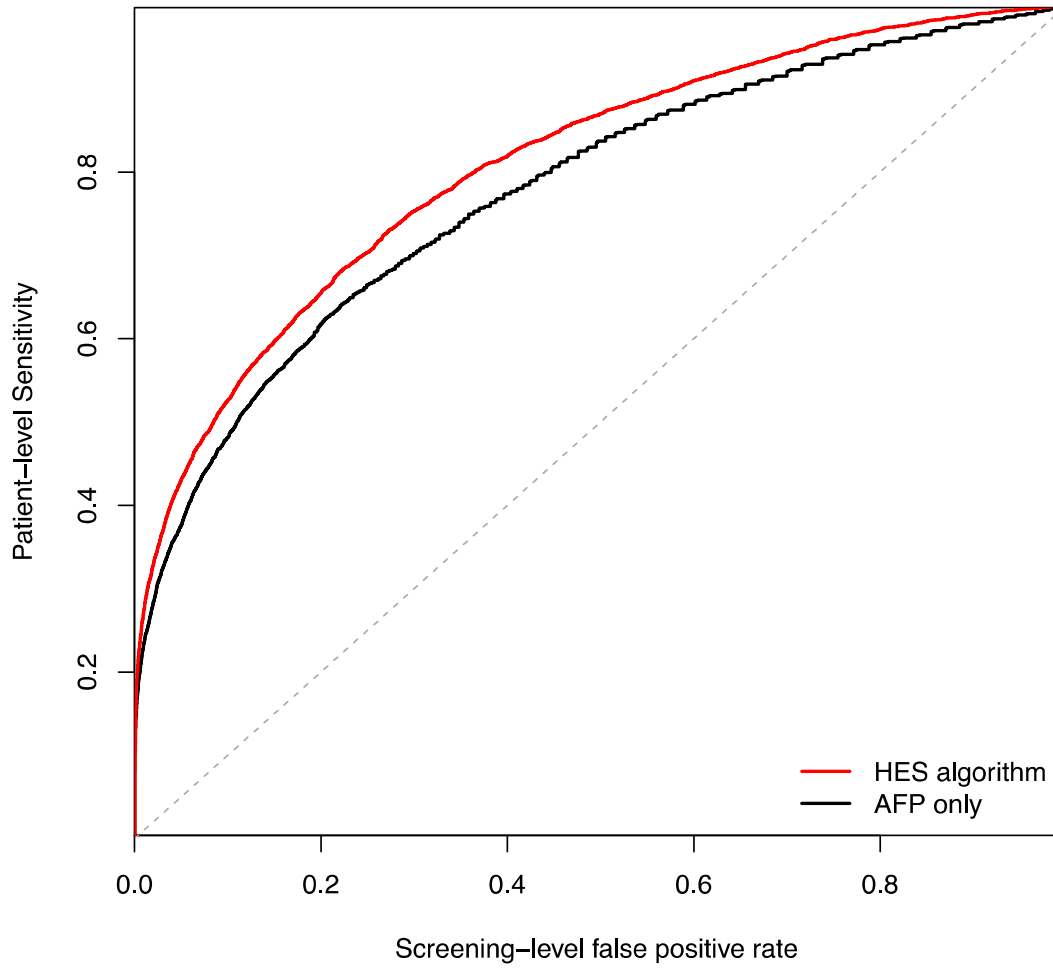
**Supplementary Figure 3:** Visual illustration of periods during which a positive screening result was considered to be a true positive screen and periods where a positive screen result was considered to be a false positive screen in HCC cases for different time-dependent definitions of patient-level sensitivity and screening-level false positive rate.
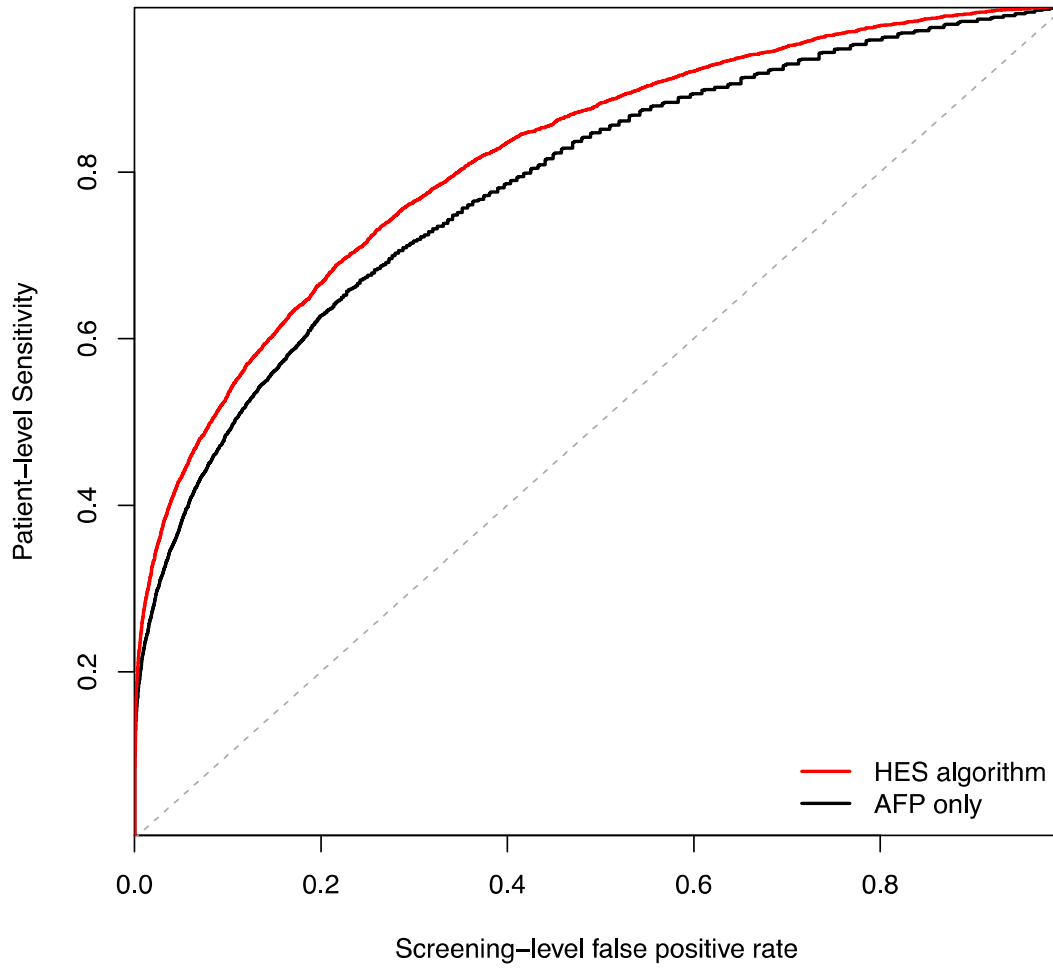
38  **Supplementary Figure 4:** Receiver operating characteristic (ROC) curves within T=6 months
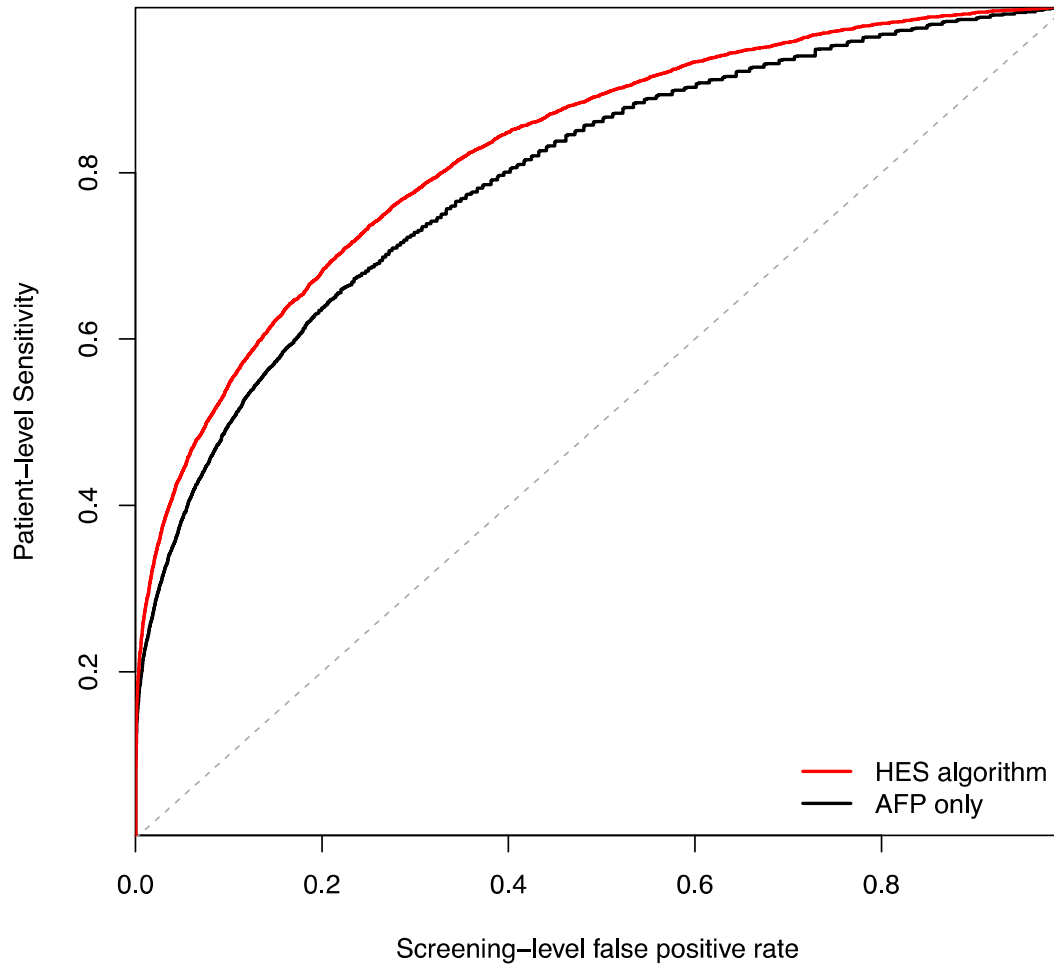39  prior to HCC diagnosis.



40

41

42  **Supplementary Figure 5:** Receiver operating characteristic (ROC) curves within T=12 months
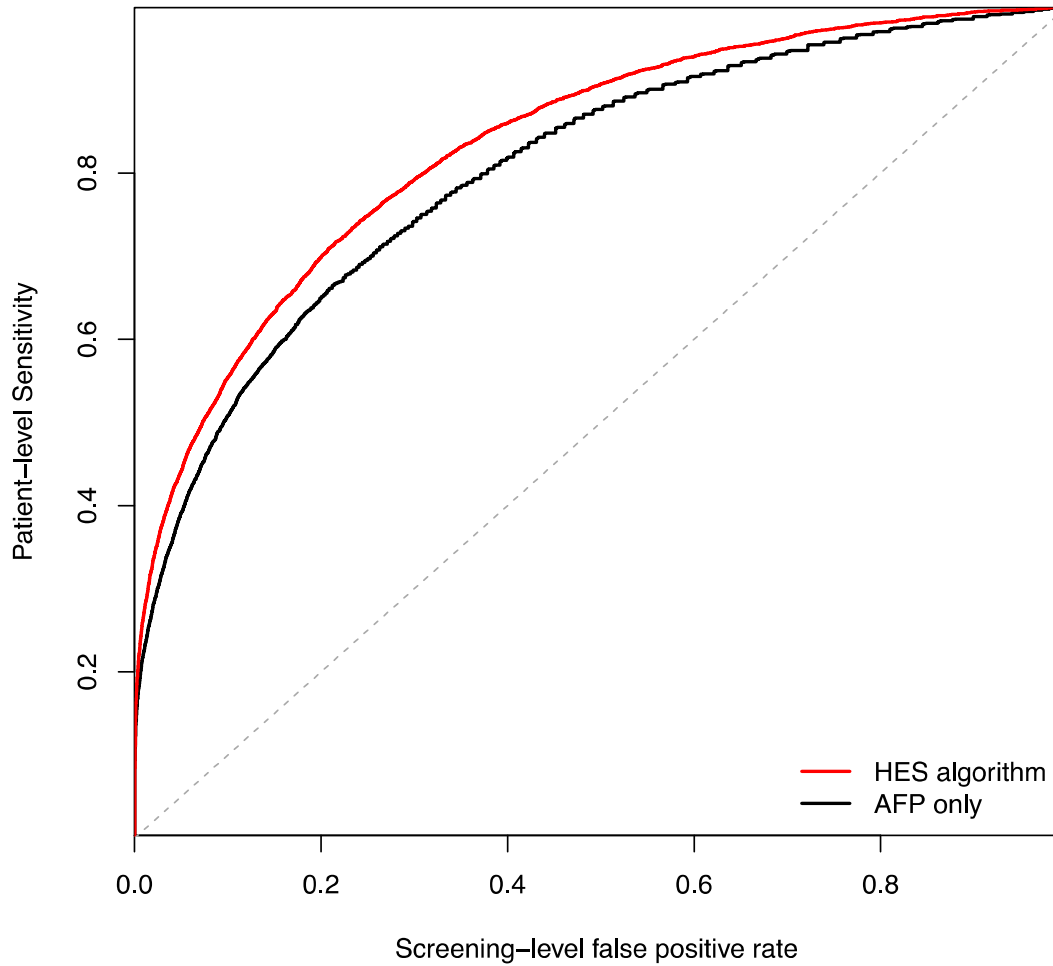43  prior to HCC diagnosis.



44

45

46

47  **Supplementary Figure 6:** Receiver operating characteristic (ROC) curves within T=24 months
48  prior to HCC diagnosis.



49

50

51    **Supplementary Figure 7:** Receiver operating characteristic (ROC) curves any time prior to HCC
52    diagnosis.



53

54     **Supplementary Table 2:** Alternative definition of HCC diagnosis time frame

| Patient-level sensitivity corresponding to 10% screening-level false positive rate | 4,804 HCC cases; 33,627 cirrhosis controls | | | |
|---|---|---|---|---|
| | **0-6 months** | **6-12 months** | **12-24 months** | **24-Ever** |
| HES algorithm | 52.56 (95% CI: 50.36-53.20) | 36.62 (95% CI: 33.52-37.79) | 28.41 (95% CI: 26.21-28.77) | 26.06 (95% CI: 23.47-27.01) |
| AFP only | 48.13 (95% CI: 46.48-48.62) | 33.87 (95% CI: 31.45-35.14) | 26.40 (95% CI: 23.99-27.28) | 25.21 (95% CI: 22.08-26.30) |
| Bootstrap p-values: | <0.0005 | <0.0005 | <0.0005 | 0.095 |

55

56

57  **Supplementary Table 3:** Evaluating the HES algorithm in etiology subgroups

| Patient-level sensitivity corresponding to 10% screening-level false positive rate | Hepatitis C only* 337 HCC cases; 1,888 cirrhosis controls | | | |
|---|---|---|---|---|
| | 6 months | 12 months | 24 months | Ever |
| HES algorithm | 42.96 (95% CI: 35.57-48.33) | 42.48 (95% CI: 36.39-47.87) | 43.16 (95% CI: 36.92-48.72) | 43.32 (95% CI: 36.51-48.83) |
| AFP only | 38.99 (95% CI: 31.90-44.19) | 37.26 (95% CI: 30.46-41.97) | 37.08 (95% CI: 29.47-41.69) | 38.28 (95% CI: 31.45-43.90) |
| Bootstrap p-values: | 0.0205 | 0.0060 | 0.0050 | 0.0050 |

58  Patients included in this analysis are hepatitis C positive, hepatitis B negative/unknown, have no
59  alcoholic liver disease and no non-alcoholic fatty liver disease.

| Patient-level sensitivity corresponding to 10% screening-level false positive rate | Alcoholic liver disease only* 945 HCC cases; 11,294 cirrhosis controls | | | |
|---|---|---|---|---|
| | 6 months | 12 months | 24 months | Ever |
| HES algorithm | 57.54 (95% CI: 53.59-60.60) | 57.71 (95% CI: 54.22-60.87) | 58.03 (95% CI: 54.20-60.83) | 59.33 (95% CI: 55.77-62.24) |
| AFP only | 53.65 (95% CI: 49.70-56.56) | 53.71 (95% CI: 50.35-56.99) | 53.99 (95% CI: 50.35-57.11) | 54.51 (95% CI: 50.94-58.08) |
| Bootstrap p-values: | <0.0005 | 0.0005 | <0.0005 | <0.0005 |

60  Patients included in this analysis have alcoholic liver disease, are hepatitis C negative/unknown,
61  hepatitis B negative/unknown, and have no non-alcoholic fatty liver disease.

| Patient-level sensitivity corresponding to 10% screening-level false positive rate | Hepatitis C or Alcoholic liver disease only* 3,737 HCC cases; 23,996 cirrhosis controls | | | |
|---|---|---|---|---|
| | 6 months | 12 months | 24 months | Ever |
| HES algorithm | 51.88 (95% CI: 49.62-52.71) | 52.39 (95% CI: 49.96-53.12) | 53.59 (95% CI: 51.78-54.23) | 54.70 (95% CI: 52.79-55.32) |
| AFP only | 47.40 (95% CI: 45.61-48.04) | 47.93 (95% CI: 45.75-48.50) | 48.92 (95% CI: 46.88-49.54) | 50.20 (95% CI: 47.77-50.97) |
| Bootstrap p-values: | <0.0005 | <0.0005 | <0.0005 | <0.0005 |

62  Patients included in this analysis are either hepatitis C positive or have alcoholic liver disease, are
63  hepatitis B negative/unknown, and no non-alcoholic fatty liver disease.