

NFL COMBINE PERFORMANCE vs. CAREER SUCCESS

A Position-by-Position Analysis of Which Drills Predict
NFL Contributor vs. Bust Outcomes

Author: Armaan Sharma
Data Source: NFL Combine 2010–2024 (nflreadr / nflverse)
Career Data: Pro Football Reference via nflreadr
Sample: 5,062 combine invitees (2010–2024); 3,634 career-complete (2010–2020 classes); 3,511 across 10 analyzed positions
Method: AUC (ROC analysis), point-biserial correlation, optimal threshold via Youden’s J
Date: June 22, 2026

This report is a quantitative supplement to film-based scouting, not a replacement. Combine metrics measure physical traits in controlled conditions; career outcomes depend on coaching quality, scheme fit, injury luck, and football intelligence that no timed drill captures.

Contents

1	Data Sources and Methodology	3
1.1	Data Sources	3
1.2	Career Outcome Tiers	3
1.3	Analytical Approach	4
2	Career Outcome Tier Distribution	5
3	AUC Heatmap: Which Drills Predict at Which Position	6
4	Position-by-Position Findings	7
5	40-Yard Dash Distributions by Tier: Skill Positions	17
6	Five Headline Cross-Position Takeaways	18
7	Limitations and Caveats	19

Data Sources and Methodology

Data Sources

All data was pulled via the `nflreadr` R package (version current as of June 22, 2026). No manual data entry; all figures are reproducible by re-running the pipeline scripts.

Source	Contents and Use
<code>load_combine()</code>	Combine results for all invited participants, 2010–2024. Includes forty-yard dash, bench press (reps), vertical jump, broad jump, 3-cone drill, and shuttle run. Also contains draft outcome fields (draft team, round, overall pick) embedded per player.
<code>load_draft_picks()</code>	Drafted players with Pro Football Reference career statistics: weighted career AV (<code>w_av</code>), seasons started, games played, Pro Bowl selections, and All-Pro selections. Note: <code>car_av</code> (career AV) is entirely missing from this dataset; <code>w_av</code> (weighted, emphasising peak seasons) is used throughout.
<code>load_players()</code>	All players who ever appeared on an NFL 53-man or practice-squad roster. Used to identify combine invitees who went undrafted but still made an NFL roster (UDFA pathway).

Career Outcome Tiers

Each player in the dataset is assigned to one of four career success tiers. The backbone metric is **seasons started** from PFR, which is position-neutral (unlike weighted AV, which systematically favors high-touch skill positions). Pro Bowl and All-Pro selections identify the elite tier.

Tier	Label	Exact Criteria
1	Did Not Contribute	Undrafted <i>and</i> not on any NFL roster; <i>or</i> drafted but 0 seasons started <i>and</i> weighted AV ≤ 10 .
2	Contributor / Rotational	Drafted with ≥ 1 season as a starter <i>or</i> weighted AV > 10 (catches high-value subs who weren't formally listed as "starter"); <i>or</i> undrafted but appeared on an NFL roster for ≥ 1 season (years-of-experience proxy).
3	Multi-Year Starter	≥ 4 seasons started (drafted); <i>or</i> undrafted UDFA with ≥ 4 years of NFL experience (rare; ~ 30 players in 15 years).
4	Elite	≥ 2 Pro Bowl selections <i>or</i> ≥ 1 All-Pro selection. Pro Bowls ≥ 2 to filter out honorary/injury-replacement appearances.

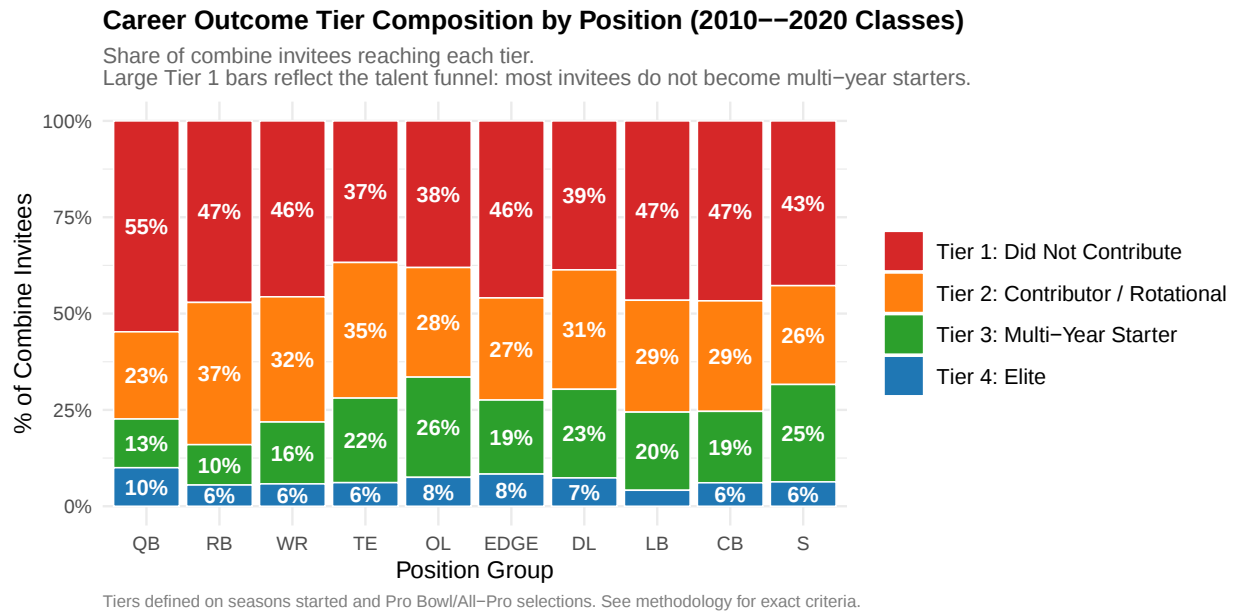
Provisional outcomes (2021–2024 classes): Players from the 2021–2024 draft classes are included in the dataset but *excluded from all predictive analysis* (AUC, correlation, threshold estimation). Their careers are too short to reliably classify into tiers 3 or 4. They are flagged in the data with `career_complete = FALSE`. All AUC and correlation statistics in this report derive from the 2010–2020 draft classes only ($n = 3,634$). Of these, 123 are specialists (64 punters, 46 kickers, 13 long snappers) who are excluded from the position-by-position findings due to small sample size and the distinct nature of their evaluation. The ten primary position groups analyzed (QB through S) account for $n = 3,511$ of the career-complete cohort.

Analytical Approach

For each position group $\times$ drill combination with $n \geq 30$ complete-career players:

1. **Tier mean comparison:** Average drill performance per tier (Tiers 1–4), with standard deviations.
2. **AUC:** Area under the ROC curve for predicting “starter-or-better” (Tier 3+) from the drill value, computed with the **pROC** package. $\text{AUC} = 0.50$ means no better than chance; $= 1.0$ means perfect separation.
3. **Correlation:** Point-biserial correlation (r) between the drill value and the binary starter-or-better outcome. For “lower-is-better” drills (40-yard, cone, shuttle), the sign is flipped so that positive r always means “more of this $\Rightarrow$ better career.”
4. **Predictive classification:** *Strong* = $\text{AUC} \geq 0.60$ and $|r| \geq 0.15$; *Moderate* = $\text{AUC} \geq 0.55$ and $|r| \geq 0.10$; *Weak/None* = below moderate thresholds.
5. **Optimal threshold:** The drill cutpoint that maximises Youden’s J statistic (sensitivity + specificity -1). Reported as the value that best separates Tier 3+ from Tier 1–2 players in the cohort.

Career Outcome Tier Distribution



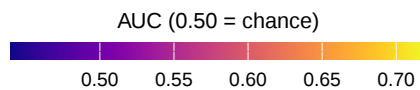
The large Tier 1 bar across all positions reflects the structure of the NFL talent funnel. Among combine invitees (already an elite pre-screened group), roughly 49% do not become meaningful contributors. Quarterbacks have the highest Tier 4 (elite) rate at 10%, reflecting the high draft capital invested in the position. Wide receivers and running backs have among the highest Tier 1 rates, consistent with their combine data showing no predictive signal; these positions have the highest variance in outcomes relative to tested athleticism.

AUC Heatmap: Which Drills Predict at Which Position

Combine Drill Predictive Power by Position: AUC

AUC for predicting Tier 3+ (multi-year starter or elite). Brighter = more predictive. 0.50 = no better than chance.

	Bench Press	40-Yard Dash	Vertical Jump	Broad Jump	3-Cone Drill	Shuttle Run
RB	0.479	0.518	0.503	0.476	0.544	0.573
WR	0.529	0.544	0.507	0.547	0.513	0.516
TE	0.561	0.609	0.568	0.558	0.671	0.624
OL	0.585	0.637	0.592	0.630	0.621	0.646
EDGE	0.524	0.607	0.553	0.560	0.581	0.583
DL	0.594	0.526	0.554	0.536	0.565	0.535
LB	0.601	0.634	0.569	0.527	0.608	0.644
CB	0.523	0.587	0.505	0.569	0.508	0.547
S	0.582	0.568	0.654	0.650	0.640	0.593
QB	n/a	0.607	0.519	0.590	0.618	0.549



Based on 2010--2020 career-complete cohort. Strong = AUC $\geq$ 0.60; Moderate = 0.55--0.59; Weak/None < 0.55.

Key pattern: The bottom half of the heatmap (OL, EDGE, DL, LB) is consistently brighter than the top half (WR, RB, CB). Line positions are more predictable by combine metrics than skill positions. The 3-cone drill column is among the brightest overall, while the bench press column is consistently among the dimmest for most positions.

Position-by-Position Findings

Quarterbacks

Career-complete cohort: n = 190 | Tier 3+ (starters): 43 (22.6%) | Tier 4 (elite): 19 (10%)

Mobility reads as durability and pocket-escape ability at the NFL level. The 40-yard dash and 3-cone drill both signal starter outcomes for quarterbacks. Every Tier 4 QB in the career-complete cohort ran sub-4.85 in the 40 and sub-7.20 on the 3-cone. Bench press data is too sparse to evaluate meaningfully. The vertical jump showed no predictive value (AUC = 0.519).

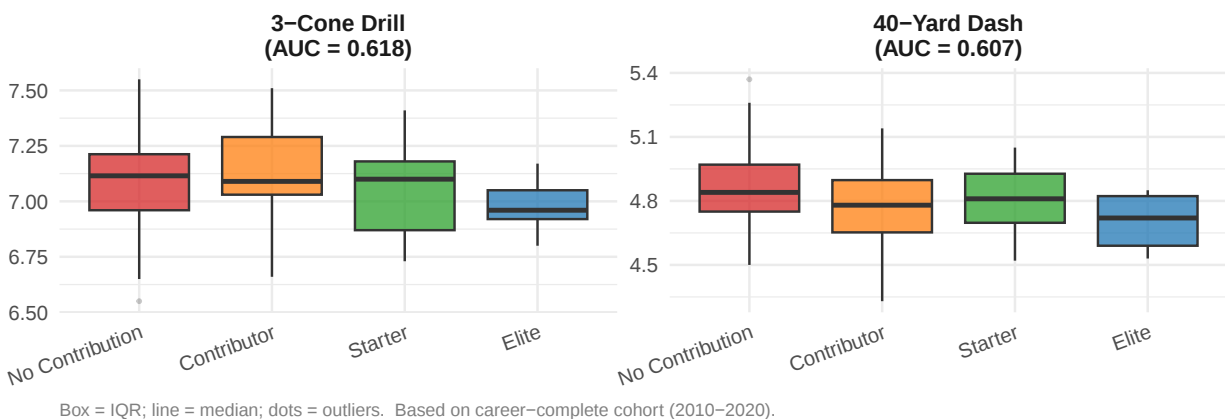
Drill	Direction	n	r	AUC	Signal	Threshold
3-Cone Drill	Lower = Better	149	+0.160	0.618	Strong	sub-7.20s
40-Yard Dash	Lower = Better	177	+0.156	0.607	Strong	sub-4.85s
Broad Jump	Higher = Better	164	+0.127	0.590	Moderate	≥116.00"
Shuttle Run	Lower = Better	152	+0.062	0.549	Weak/None	sub-4.32s
Vertical Jump	Higher = Better	163	+0.016	0.519	Weak/None	≥32.50"

RELY ON

3-Cone Drill, 40-Yard Dash

STOP OVER-VALUING

Shuttle Run, Vertical Jump



Running Backs

Career-complete cohort: n = 363 | Tier 3+ (starters): 58 (16%) | Tier 4 (elite): 20 (5.5%)

This is the most counterintuitive result in the entire dataset: no combine drill predicts running back career success above the moderate threshold. The 40-yard dash (the single number most associated with RB evaluation) produces $AUC = 0.518$ (essentially a coin flip). Bench press (0.479) is actually slightly below chance. The shuttle run is the closest to a signal (0.573) but doesn't clear the moderate bar. Within the already-elite pool of combine invitees, RB evaluation must lean entirely on film, vision, pass-protection technique, and football intelligence; traits no timed drill measures.

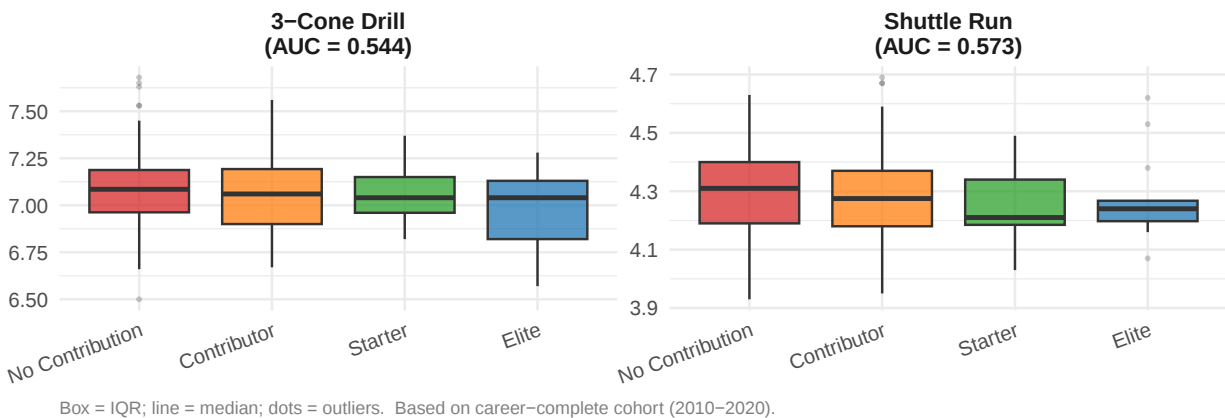
Drill	Direction	n	r	AUC	Signal	Threshold
Shuttle Run	Lower = Better	225	+0.088	0.573	Weak/None	sub-4.27s
3-Cone Drill	Lower = Better	212	+0.074	0.544	Weak/None	sub-7.15s
40-Yard Dash	Lower = Better	346	-0.018	0.518	Weak/None	sub-4.73s
Vertical Jump	Higher = Better	316	+0.009	0.503	Weak/None	$\geq 37.50"$
Bench Press	Higher = Better	304	-0.028	0.479	Weak/None	≥ 15.00 reps
Broad Jump	Higher = Better	305	+0.031	0.476	Weak/None	$\geq 116.00"$

RELY ON

No drill reaches strong threshold for RB.

STOP OVER-VALUING

Shuttle Run, 3-Cone Drill, 40-Yard Dash, Vertical Jump, Bench Press, Broad Jump



Wide Receivers

Career-complete cohort: n = 517 | Tier 3+ (starters): 113 (21.9%) | Tier 4 (elite): 30 (5.8%)

Like running backs, wide receivers show no predictive signal from any combine drill. 40-yard dash AUC = 0.544, bench press = 0.529, vertical = 0.507, broad jump = 0.547, 3-cone = 0.513, shuttle = 0.516. The received wisdom that 4.3-speed separates elite receivers is not supported in this data. The compressed range of athleticism among combine invitees means the remaining variance in speed does not predict outcomes. Route precision, separation technique, contested catch ability, and release quickness are the separating factors, and the combine doesn't test any of them.

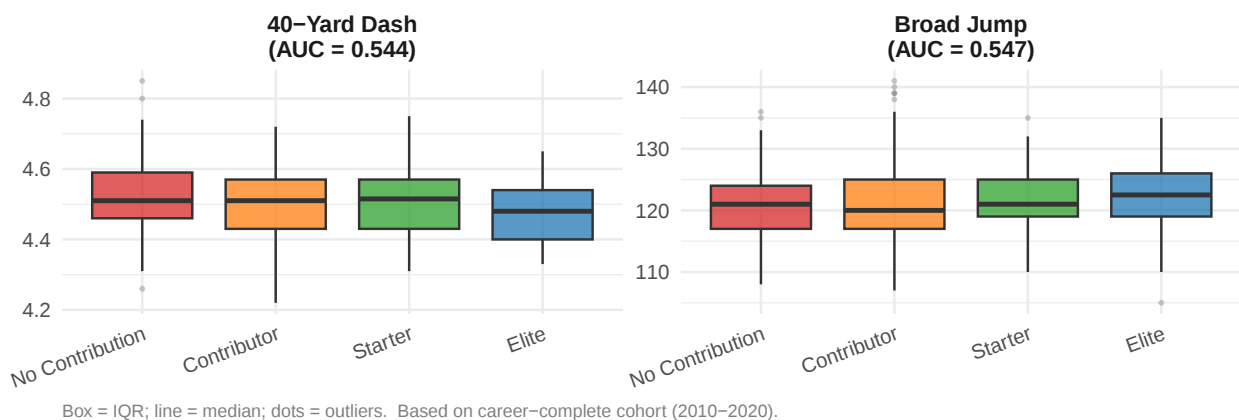
Drill	Direction	n	r	AUC	Signal	Threshold
Broad Jump	Higher = Better	439	+0.035	0.547	Weak/None	≥119.00"
40-Yard Dash	Lower = Better	481	+0.072	0.544	Weak/None	sub-4.43s
Bench Press	Higher = Better	402	+0.034	0.529	Weak/None	≥14.00 reps
Shuttle Run	Lower = Better	351	+0.039	0.516	Weak/None	sub-4.39s
3-Cone Drill	Lower = Better	335	-0.014	0.513	Weak/None	sub-7.09s
Vertical Jump	Higher = Better	446	-0.001	0.507	Weak/None	≥35.00"

RELY ON

No drill reaches strong threshold for WR.

STOP OVER-VALUING

Broad Jump, 40-Yard Dash, Bench Press, Shuttle Run, 3-Cone Drill, Vertical Jump



Tight Ends

Career-complete cohort: n = 196 | Tier 3+ (starters): 55 (28.1%) | Tier 4 (elite): 12 (6.1%)

Tight ends are the most clearly predicted position by agility, with the 3-cone drill as the best single predictor in the entire dataset (AUC = 0.671). A TE who can bend and change direction translates to NFL route-running as a receiver; one who can only run straight lines does not. The shuttle run is also strong (0.624), and the 40-yard dash is meaningful (0.609). Notably, bench press is the weakest predictor (0.561, barely moderate); physical dominance at the line does not predict whether a TE becomes a multi-year starter.

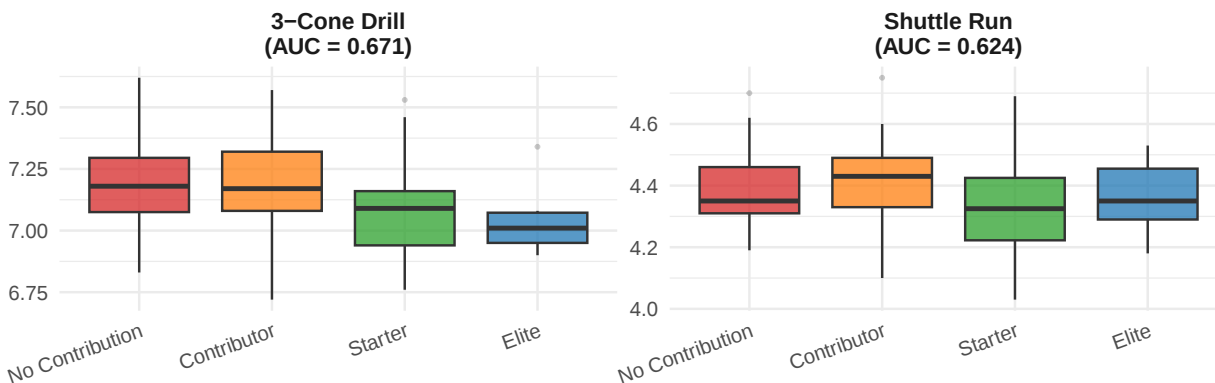
Drill	Direction	n	r	AUC	Signal	Threshold
3-Cone Drill	Lower = Better	125	+0.247	0.671	Strong	sub-7.13s
Shuttle Run	Lower = Better	126	+0.213	0.624	Strong	sub-4.23s
40-Yard Dash	Lower = Better	181	+0.161	0.609	Strong	sub-4.68s
Vertical Jump	Higher = Better	146	+0.106	0.568	Moderate	≥37.00"
Bench Press	Higher = Better	152	+0.085	0.561	Weak/None	≥23.00 reps
Broad Jump	Higher = Better	141	+0.109	0.558	Moderate	≥119.00"

RELY ON

3-Cone Drill, Shuttle Run, 40-Yard Dash

STOP OVER-VALUING

Bench Press



Offensive Linemen

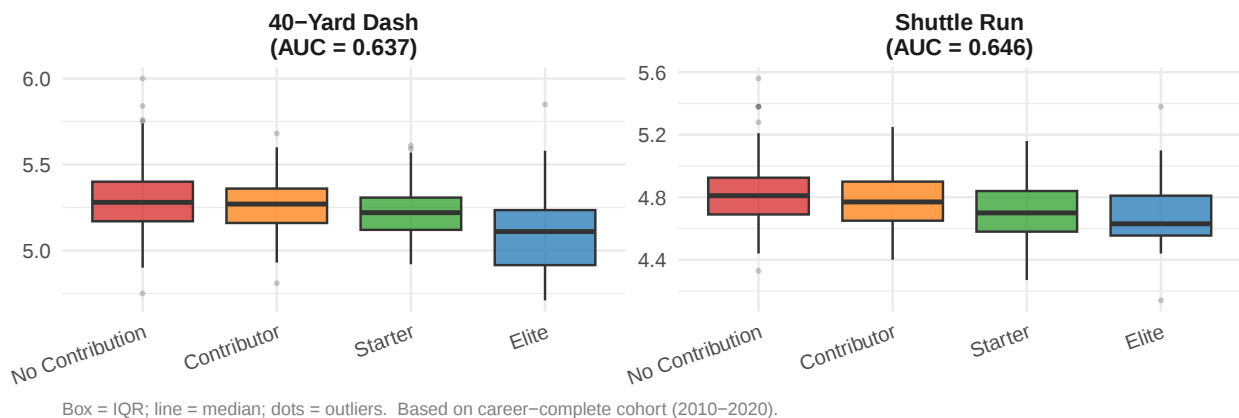
Career-complete cohort: n = 573 | Tier 3+ (starters): 192 (33.5%) | Tier 4 (elite): 43 (7.5%)

Offensive linemen are the most consistently predicted position group in the dataset, with four drills reaching the strong threshold. This is counterintuitive (we tend to think of linemen as less athletically differentiated) but the data shows that movement athleticism (not raw strength) is what separates starters from busts. The shuttle run (0.646) and 40-yard dash (0.637) are both strong predictors for big men. Linemen who can move laterally, change direction, and get off the snap quickly tend to start in the NFL; stiff, slow linemen don't.

Drill	Direction	n	r	AUC	Signal	Threshold
Shuttle Run	Lower = Better	426	+0.244	0.646	Strong	sub-4.64s
40-Yard Dash	Lower = Better	524	+0.227	0.637	Strong	sub-5.25s
Broad Jump	Higher = Better	442	+0.216	0.630	Strong	≥103.00"
3-Cone Drill	Lower = Better	412	+0.219	0.621	Strong	sub-7.70s
Vertical Jump	Higher = Better	453	+0.156	0.592	Moderate	≥26.00"
Bench Press	Higher = Better	463	+0.121	0.585	Moderate	≥26.00 reps

RELY ON

Shuttle Run, 40-Yard Dash, Broad Jump, 3-Cone Drill



Edge Rushers (EDGE/DE/OLB)

Career-complete cohort: n = 479 | Tier 3+ (starters): 132 (27.6%) | Tier 4 (elite): 40 (8.4%)

Edge rushers are predicted primarily by first-step speed (40-yard dash, AUC = 0.607, strong). The 3-cone (0.581) and shuttle (0.583) provide moderate additional signal. Bench press is weak (0.524); this is the result that most directly contradicts a common scouting belief. Raw upper-body strength at the combine does not differentiate starter-quality edge rushers from busts; explosive speed and change-of-direction do.

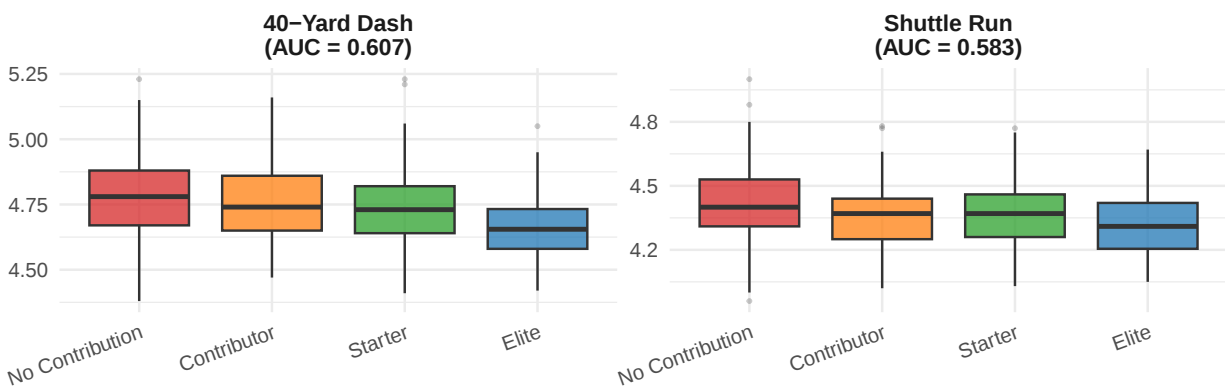
Drill	Direction	n	r	AUC	Signal	Threshold
40-Yard Dash	Lower = Better	461	+0.152	0.607	Strong	sub-4.75s
Shuttle Run	Lower = Better	330	+0.135	0.583	Moderate	sub-4.24s
3-Cone Drill	Lower = Better	327	+0.144	0.581	Moderate	sub-7.20s
Broad Jump	Higher = Better	394	+0.112	0.560	Moderate	≥115.00"
Vertical Jump	Higher = Better	391	+0.085	0.553	Weak/None	≥33.50"
Bench Press	Higher = Better	372	+0.054	0.524	Weak/None	≥28.00 reps

RELY ON

40-Yard Dash

STOP OVER-VALUING

Vertical Jump, Bench Press



Interior Defensive Linemen

Career-complete cohort: n = 313 | Tier 3+ (starters): 95 (30.4%) | Tier 4 (elite): 23 (7.3%)

Interior defensive linemen are the only position where bench press is the top predictor (AUC = 0.594, moderate). This makes functional sense: a one-gap penetrator or 0-technique nose tackle needs to resist and absorb double teams, which requires functional strength. The 3-cone (0.565) and vertical jump (0.554) provide additional moderate signal. The 40-yard dash (obsessively clocked for DTs at the combine) is the weakest predictor (0.526). A fast 40 time for an interior lineman does not translate to NFL starter outcomes.

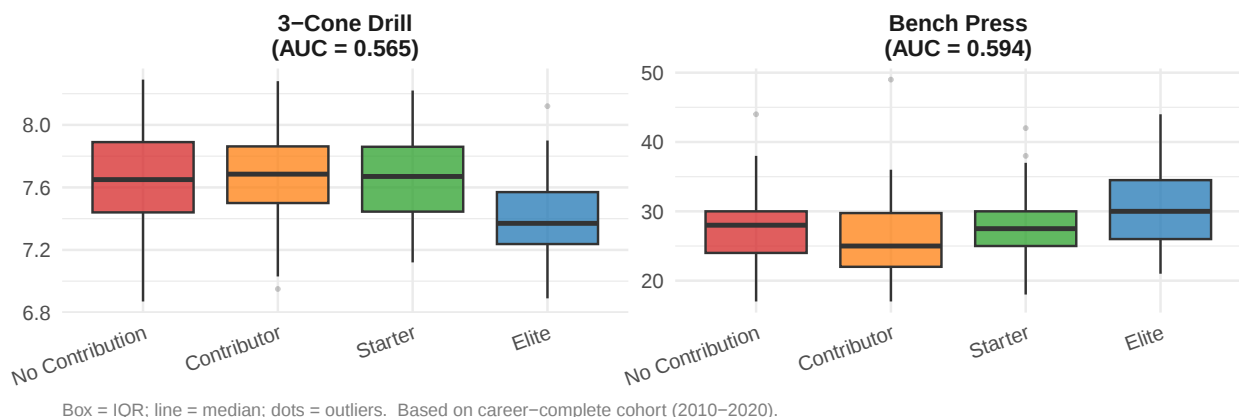
Drill	Direction	n	r	AUC	Signal	Threshold
Bench Press	Higher = Better	248	+0.152	0.594	Moderate	≥24.00 reps
3-Cone Drill	Lower = Better	214	+0.112	0.565	Moderate	sub-7.39s
Vertical Jump	Higher = Better	244	+0.111	0.554	Moderate	≥29.50"
Broad Jump	Higher = Better	236	+0.069	0.536	Weak/None	≥101.00"
Shuttle Run	Lower = Better	219	+0.073	0.535	Weak/None	sub-4.75s
40-Yard Dash	Lower = Better	293	+0.061	0.526	Weak/None	sub-5.11s

RELY ON

No drill reaches strong threshold for DL.

STOP OVER-VALUING

Broad Jump, Shuttle Run, 40-Yard Dash



Linebackers (LB/ILB)

Career-complete cohort: n = 217 | Tier 3+ (starters): 53 (24.4%) | Tier 4 (elite): 9 (4.1%)

True inside linebackers (ILB/LB in 4-3 and 3-4 schemes) are the most ‘combineable’ position in the dataset: four separate drills hit the strong predictive threshold simultaneously. This suggests LB success requires a complete athletic profile: the position demands that you run, flow, shed, and cover. Any one deficiency matters. The shuttle run (0.644) is the single best predictor, with the 40-yard dash (0.634), 3-cone (0.608), and bench press (0.601) close behind. A complete athlete at LB is predictably better than a one-dimensional one.

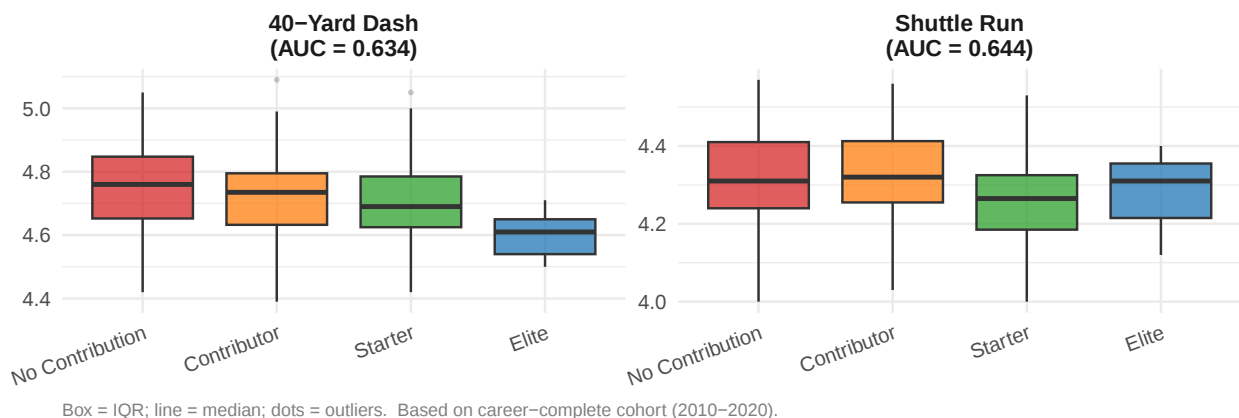
Drill	Direction	n	r	AUC	Signal	Threshold
Shuttle Run	Lower = Better	118	+0.217	0.644	Strong	sub-4.34s
40-Yard Dash	Lower = Better	177	+0.196	0.634	Strong	sub-4.71s
3-Cone Drill	Lower = Better	107	+0.167	0.608	Strong	sub-7.00s
Bench Press	Higher = Better	165	+0.155	0.601	Strong	≥22.00 reps
Vertical Jump	Higher = Better	163	+0.123	0.569	Moderate	≥36.50"
Broad Jump	Higher = Better	164	+0.059	0.527	Weak/None	≥113.00"

RELY ON

Shuttle Run, 40-Yard Dash, 3-Cone Drill, Bench Press

STOP OVER-VALUING

Broad Jump



Cornerbacks

Career-complete cohort: $n = 394$ | Tier 3+ (starters): 97 (24.6%) | Tier 4 (elite): 24 (6.1%)

Cornerback evaluation reveals interesting limits of combine metrics. Speed (40-yard dash, $AUC = 0.587$) and lateral explosiveness (broad jump, 0.569) both show moderate predictive signal. But the 3-cone and shuttle (arguably more relevant for the actual movements of man coverage) are weak predictors. One explanation: the best CBs in each draft class get selected regardless of agility scores, compressing the observable range in those metrics. The 40-yard dash, by contrast, has more variance preserved at the top of the draft pool.

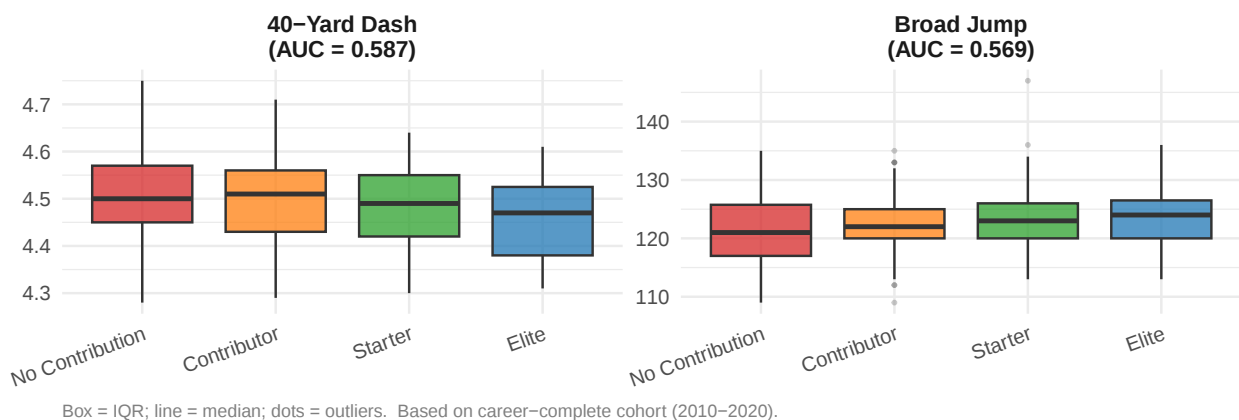
Drill	Direction	n	r	AUC	Signal	Threshold
40-Yard Dash	Lower = Better	377	+0.144	0.587	Moderate	sub-4.43s
Broad Jump	Higher = Better	320	+0.129	0.569	Moderate	$\geq 123.00"$
Shuttle Run	Lower = Better	246	+0.086	0.547	Weak/None	sub-4.15s
Bench Press	Higher = Better	311	+0.029	0.523	Weak/None	≥ 14.00 reps
3-Cone Drill	Lower = Better	240	-0.007	0.508	Weak/None	sub-7.04s
Vertical Jump	Higher = Better	319	+0.029	0.505	Weak/None	$\geq 37.50"$

RELY ON

No drill reaches strong threshold for CB.

STOP OVER-VALUING

Shuttle Run, Bench Press, 3-Cone Drill, Vertical Jump



Safeties

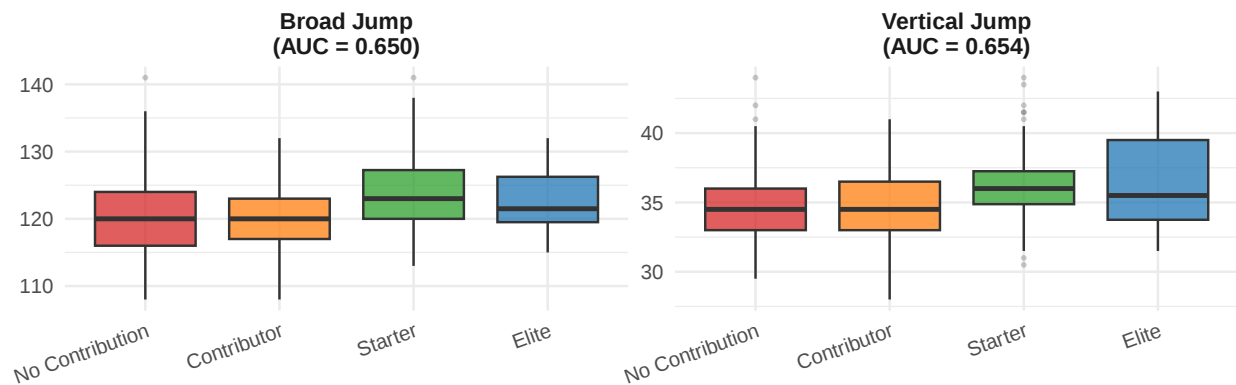
Career-complete cohort: n = 269 | Tier 3+ (starters): 85 (31.6%) | Tier 4 (elite): 17 (6.3%)

Safeties are predicted more by explosiveness than straight-line speed. Vertical jump (AUC = 0.654) and broad jump (0.650) are both strong predictors and outperform the 40-yard dash (0.568, moderate). The 3-cone drill is also strong (0.640). This fits the modern safety role: burst and closing speed on short breaks, range to play centerfield, and the ability to drive on the ball. A safety's job isn't to run 40 yards in a straight line; it's to close on a receiver 10 yards away or rotate over the top at the snap.

Drill	Direction	n	r	AUC	Signal	Threshold
Vertical Jump	Higher = Better	218	+0.254	0.654	Strong	≥35.00"
Broad Jump	Higher = Better	218	+0.238	0.650	Strong	≥121.00"
3-Cone Drill	Lower = Better	143	+0.213	0.640	Strong	sub-6.86s
Shuttle Run	Lower = Better	152	+0.137	0.593	Moderate	sub-4.20s
Bench Press	Higher = Better	213	+0.100	0.582	Moderate	≥16.00 reps
40-Yard Dash	Lower = Better	239	+0.101	0.568	Moderate	sub-4.56s

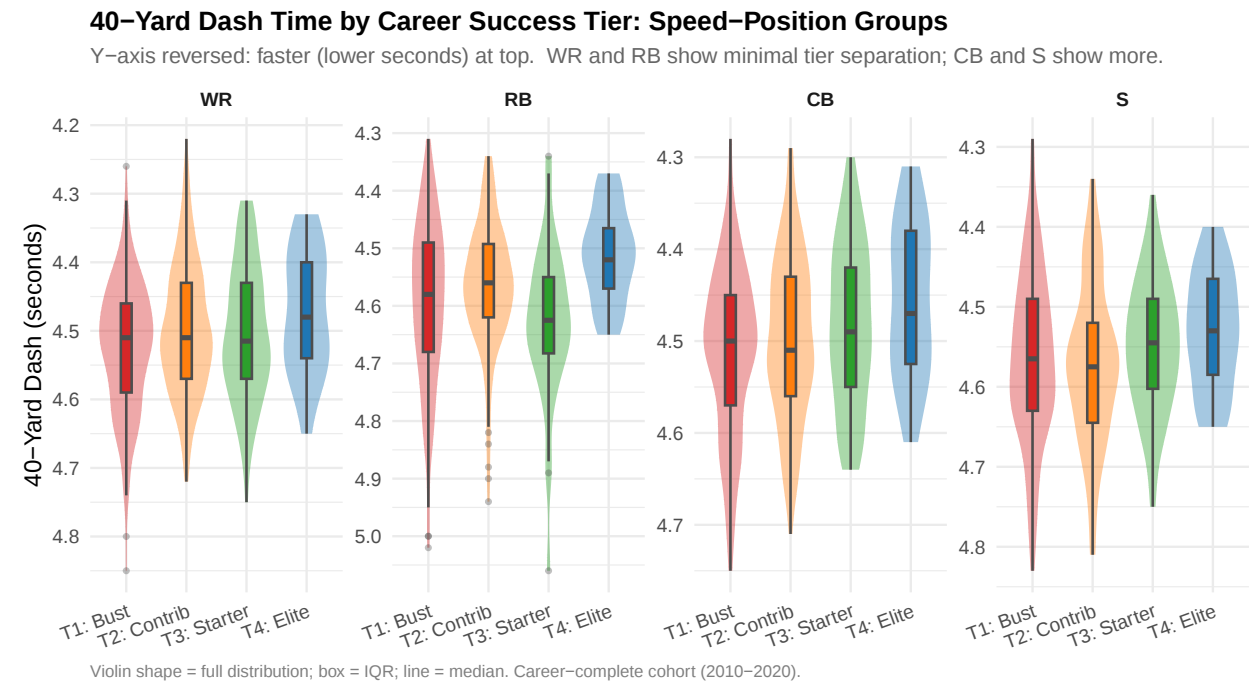
RELY ON

Vertical Jump, Broad Jump, 3-Cone Drill



Box = IQR; line = median; dots = outliers. Based on career-complete cohort (2010–2020).

40-Yard Dash Distributions by Tier: Skill Positions



The most striking feature of this chart is how similar the four tier distributions are for wide receivers and running backs. The violin shapes overlap almost entirely across Tier 1 through Tier 4 for both positions, consistent with the near-zero predictive AUC values. Cornerbacks and safeties show slightly more tier separation; the Tier 4 elites tend to cluster toward faster times, but the overlap remains substantial.

Five Headline Cross-Position Takeaways

1. WR and RB Combine Numbers Are Noise

No combine drill reaches even a moderate predictive threshold for wide receivers or running backs. The 40-yard dash (the single most-discussed number at the combine) produces $AUC = 0.544$ for WRs and 0.518 for RBs, both statistically indistinguishable from random. Teams that overweight combine results for these positions are optimising the wrong signal. Film-based evaluation of route technique, vision, hands, and pass-protection must dominate.

2. The 3-Cone Drill Is the Most Underrated Test

The 3-cone drill is the strongest single predictor in the entire dataset (TE: $AUC = 0.671$). When considering both agility drills together, either the 3-cone or the shuttle run outperforms the 40-yard dash at 7 of 10 positions (QB, RB, TE, OL, DL, LB, S). The 3-cone alone beats the 40 at 5 positions (QB, RB, TE, DL, S). Despite this, the combine media cycle obsesses over 40-yard dash times and barely mentions cone or shuttle results. For OL, S, LB, QB, and TE evaluation, agility drills deserve significantly more weight than they currently receive. Football-specific change-of-direction, not straight-line speed, is what separates starters from busts at most positions.

3. OL Athleticism Is Fully Measurable and Fully Predictive

Offensive linemen are the position most consistently predicted by combine athleticism, with four movement drills (shuttle: $AUC = 0.646$; 40-yard: 0.637 ; broad jump: 0.630 ; 3-cone: 0.621) all reaching the strong threshold. The sample here is the largest in the study ($n = 573$ complete-career linemen), giving high confidence in these findings. The “athlete OL” narrative is not hype: linemen who can move laterally, get to the second level, and redirect tend to start in the NFL; stiff, slow linemen don’t, regardless of how they look on tape physically.

4. True Linebackers Are the Most Combine-Predictable Position

Four separate drills reach strong predictive thresholds for inside linebackers, more than any other position group. The shuttle run ($AUC = 0.644$), 40-yard dash (0.634), 3-cone drill (0.608), and bench press (0.601) all carry meaningful signal. This breadth suggests that LB success requires a complete athletic profile: speed to flow to the ball, strength to shed blocks, agility to cover in space. Any single athletic deficiency at the position tends to show up in outcomes.

5. The 40-Yard Dash Predicts Line Play, Not Skill-Position Play

The 40-yard dash is a strong predictor for OL (0.637), LB (0.634), EDGE (0.607), TE (0.609), and QB (0.607). It is moderate for CB (0.587) and S (0.568). It is essentially random for WR (0.544) and RB (0.518). Speed predicts *linemen and hybrid pass-catchers* more than it predicts traditional skill-position outcomes, the opposite of how 40 times are discussed in draft coverage.

Limitations and Caveats

1. Survivorship and Selection Bias. Combine invitees represent the top 1–2% of college players who attracted sufficient NFL interest to receive an invitation. The variance in athleticism within this pool is compressed relative to the full population of college players. A low AUC for a drill at a given position does not mean “athleticism doesn’t matter for that position in football.” It means that within the already-athletic pool of combine invitees, the remaining spread in that drill does not reliably separate future starters from non-contributors. Speed clearly matters for wide receivers in football; the finding here is that the between-player speed variation *among combine invitees* doesn’t predict outcomes well.

2. Missing Drill Data. Players routinely skip drills due to injury, agent advice, or personal choice. Miss rates vary significantly: bench press (36% missing), 3-cone (43% missing), shuttle (41% missing), vertical jump (22% missing), broad jump (23% missing), 40-yard dash (13% missing). Players who skip a drill are systematically different from those who participate; elite WR prospects often skip the bench press, knowing their value isn’t strength-based. This introduces selection bias in drill-specific samples that cannot be fully corrected.

3. Incomplete Careers for 2021–2024 Classes. Approximately 1,400 combine participants from the 2021–2024 draft classes have had at most 4 NFL seasons. Their outcome tiers are provisional: many Tier 1 players from these classes may still develop, and Tier 2 players may ascend to Tier 3 as their careers progress. These players are excluded from all predictive analysis but would shift results (likely toward the null) if included with partial career data.

4. Career Metric Limitations. The primary career outcome metric, seasons started from Pro Football Reference, is imperfect. “Starter” designation depends on team decisions that may reflect team-level factors (injuries, coaching preferences, scheme fit) as much as player quality. Weighted career AV (`w_av`), used as a secondary check, is known to favour high-touch positions (QB, RB, WR) and undervalue elite linemen and interior defenders. For undrafted UDFA players, `years_of_experience` is used as a rough proxy for career quality; this is a significant approximation; those players’ tier assignments may be incorrect in either direction.

5. Correlation, Not Causation. These are predictive correlations within a specific historical sample under specific NFL conditions (rule environment, team philosophies, scheme trends of 2010–2024). They do not imply that a team can improve a player’s career by training him to run a faster 40 or do more bench reps. The drill is measuring an underlying trait (acceleration, upper-body power, etc.) that is causally downstream of more fundamental physical attributes. Context entirely absent from this analysis (draft capital invested, coaching quality, team scheme fit, injury luck, off-field factors) drives the majority of variance in career outcomes.

Data pipeline: R scripts 01_pull_data.R through 05_write_report.R in the project root. All results reproducible from scratch by running scripts in order. Raw data saved to `data/raw/`; processed data to `data/processed/`; all output to `output/`.

R packages used: `nflreadr`, `dplyr`, `tidyr`, `purrr`, `ggplot2`, `pROC`, `viridis`, `ggrepel`, `cowplot`, `scales`, `knitr`, `kableExtra`, `tinytex`.