Where U.S. Men's National Soccer Team Players Come From: Production Relative to Male Youth Population (2010–2026)

Aadit Manyem

Abstract

Prior analyses have shown that elite U.S. men's soccer players come disproportionately from certain parts of the country. Building on this work, the present study provides a nationwide analysis of U.S. Men's National Team (USMNT) player production relative to the male youth population. The analysis included all 215 players who (a) were recorded playing for the USMNT from 2010 through July 3, 2026, and (b) met the criteria to be assigned a U.S. formative location, specified as city and state, using a rules-based biographical coding procedure. Each player's formative location was mapped to its commuting zone (CZ), an established geographic unit; 588 CZs partition the 50 states and D.C. without gaps or overlap. The study calculated each CZ's share of the 215 players in the sample and compared it to its share of the male youth population (U.S. males aged 5–19, 2024 data). The six CZs producing the most players accounted for 32.1% of the 215 players but only 13.7% of the male youth population. Among CZs meeting ranking inclusion criteria, those named for El Paso, Germantown, and San Diego exceeded the player production expected from their share of the male youth population by the greatest proportions; those named for Houston and Boston fell furthest below those expectations in absolute terms. These findings systematically document geographic imbalance in USMNT player production and identify areas warranting closer investigation of local player-development conditions, without establishing the observed pattern's causes.

DOI: 10.5195/pur.2026.191

Introduction

Previous analyses in journalistic outlets have documented differences across U.S. areas in elite American men's soccer player production relative to population (Hill 2019; Jackman 2017). These accounts, however, predate the last several national-team cycles, compare player output against total rather than youth population, and do not provide a common basis for ranking areas that produced no players. Building on their approach, the present study examines geographic origin patterns systematically among players who appeared for the U.S. Men's National Team (USMNT) from 2010 through July 3, 2026.

The analysis included all 215 USMNT players from this period who met the criteria to be assigned a U.S. formative location, specified as city and state, using a rules-based biographical procedure. Each player's formative location was then mapped to its commuting zone (CZ), an established geographic unit (U.S. Department of Agriculture, Economic Research Service [USDA ERS] 2026); 588 CZs collectively cover the 50 states and D.C. without gaps or overlap. The study calculated each CZ's share of the 215 players in the sample and compared it to its share of the male youth population (U.S. males aged 5–19; U.S. Census Bureau 2024). The latter population share was further used to calculate how many players each CZ would be expected to produce (a CZ containing 2% of that male youth population would be expected to produce 2% of the player sample). Among CZs meeting ranking inclusion criteria, the study ranks the ten that exceeded expectations by the greatest proportions and the ten that fell furthest below expectations. The latter ranking uses absolute differences rather than ratios because ratios cannot distinguish between CZs with zero players.

The results show that USMNT player production is highly concentrated: the six CZs producing the most players accounted for 32.1% of the player sample, despite containing just 13.7% of the male youth population. Observed player counts approached five times expected levels in some CZs, while the largest gap below expectations exceeded four players.

This analysis does not identify the causal factors underlying these geographic patterns; however, mapping them can guide future investigation and inform discussions of access to talent development and resource allocation. This is particularly relevant given documented income-based disparities in youth sport participation nationally (Solomon 2025).

Literature Review

Within men's soccer, players' geographic origins have been benchmarked against population baselines in other countries: players who represented Italy at either the youth or senior national-team level have been compared against regional population norms (Morganti et al. 2023), and Dutch elite players' regional origins have been analyzed relative to municipal youth populations (van Nieuwstadt et al. 2021). For the United States, the closest population-relative analyses of men's player origins have appeared as journalistic analyses published in soccer media.

In a feature for *American Soccer Now*, Hill (2019) assigned professionals active in fall 2018 to Census metropolitan and micropolitan areas based on where each player grew up, then grouped the areas by population size and compared their productivity using a population-per-professional ratio. Earlier,

writing for *The18*, Jackman (2017) grouped a defined sample of USMNT players by childhood region and compared each region's share of players with its share of the national population. Together, these analyses showed that professional and national-team soccer players emerge unevenly across the country and illustrated the value of comparing player origins with population.

However, five aspects of a contemporary, population-relative account of the USMNT remain unaddressed.

- Neither analysis covers the national-team period examined here. Both predate 2020, and Hill's sample was organized around professional activity rather than national-team appearance. The present study instead covers every player who appeared for the USMNT during a single recent period.
- Both analyses compared player counts against total population rather than an age- and sex-specific reference group. The present study uses a male youth baseline more closely aligned with the pool from which players are drawn.
- Neither analysis provides a common ranking of all areas producing fewer players than expected given their population sizes. Hill's population-per-professional ratio is undefined for areas producing none, while Jackman provides no corresponding ranking. The present study addresses this limitation by ranking below-expectation CZs according to how far they fall below expectation, including ones that produced no players.
- Both analyses used geographic frameworks that exclude parts of the country. Hill's metropolitan and micropolitan framework omits counties outside those classifications, while Jackman's analysis focuses on selected regions rather than a nationwide partition. The present study instead uses a nationwide geography that assigns every U.S. county to one CZ.
- Neither study specifies a reproducible procedure for assigning players to geographic origins. The present study applies a rules-based coding procedure and documents the rationale for each player's assignment, allowing another researcher to independently reproduce it.

The following section details the methods used to address these gaps.

Methodology

Sample and Dataset Construction

The present analysis examines players who appeared for the USMNT from 2010 to 2026, with the 2026 records reflecting appearances recorded as of July 3, 2026.

For 2010–2020 and 2022–2026, player lists were drawn from the official USMNT annual statistics, which report only players who recorded playing time during the relevant calendar year (U.S. Soccer n.d.-b). For 2021, because a corresponding statistics page was unavailable at the time of data collection, the player list was reconstructed from the official "2021 USMNT Lineups" page, which records the starters and substitutes used in each match (U.S. Soccer n.d.-a). These player lists were consolidated into the USMNT Formative Location Dataset, in which players appearing across multiple annual pages are represented only once. The dataset includes the following fields:

- Player Name
- Formative City
- Formative State
- Formative Location Justification
- Formative Location Supporting Source 1
- Formative Location Supporting Source 2 (if applicable)

Formative location, the constructed variable on which the analysis is based, is defined and coded as described in the following section. Because the geographic analysis requires assignment to a specific U.S. city, players classified as International (n = 38) or coded only at the state level (n = 1) were excluded, yielding an analytic sample of 215 players.

Formative Location Coding

Because no single repository documents the developmental histories of all 254 players in the dataset, sources were gathered player by player. For each player, biographical material on places of residence and transitions was located through general web searches using one or more sources that credibly established the relevant facts. Consequently, the sources vary in authoritativeness from institutional sources to journalistic and informal ones.

For the purposes of this study, a residence or transition is *documented* when it is recorded in at least one consulted source. The term therefore does not capture any transition that may have occurred but is absent from the available documentation.

The following procedure was used to assign formative locations consistently across all players. Its application to each player is recorded in the Formative Location Justification field, which contains both the coding rationale and source citation, while the Formative Location Supporting Source fields provide the corresponding links.

Identify the player's documented locations. Sources were used to organize each player's documented locations chronologically. Direct player location evidence included a residence, home, hometown, place of upbringing, or place the player was from. Club or academy participation alone was not treated as direct location evidence. When no direct city was documented, a pre-transition school city or birthplace could be used as a proxy. If no direct or proxy location was documented, a city explicitly associated with the player's sustained pre-transition youth soccer activity could be used as an activity proxy. For school city and activity proxies, a city could be drawn from a school or team name when the geographic reference was explicit. All proxy use was identified in the coding rationales.

Identify the earliest qualifying soccer-related transition. Qualifying transitions were (1) entry into a residential soccer program; (2) enrollment in a college soccer program; (3) entry into professional soccer; or (4) a documented move or residential placement to join a club or academy, attend a prep school, or otherwise obtain sustained soccer training or competition. A move under category (4) qualified only when the player's own soccer development was a documented reason for it. Entry into a residential program counted even when the player's biography did not separately mention a move. Documented entry into a full-time academy outside the player's country of documented pre-transition location was treated as

a residential placement even when the source did not separately state that the player moved. Club or academy participation without a documented move, commuting, and short-term activities did not qualify.

Assign the formative location. The formative location was the most recent directly documented player location before the earliest qualifying transition. When a non-soccer-related childhood move established a later documented residence before the earliest qualifying transition, that residence was used as the formative location. When no direct city was available, the best-supported pre-transition proxy was used. A transition occurring in the same city left that city as the formative location. Records supported only at the state level were coded as state-only; players whose documented formative development occurred outside the United States were coded as International. For International cases, documented pre-transition development within a foreign country's soccer system was sufficient for an International classification even when no foreign residence city was documented.

Taken together, these rules approximate where a player was based during development, up to the point at which soccer itself began to determine where he lived. They accommodate players who relocated shortly after birth, and they prevent later soccer-driven destinations, such as residential academies and soccer-focused prep schools, from being coded as formative since enrollment at such destinations is typically conditional on ability already demonstrated. The resulting measure is nonetheless an operational approximation and may not identify the location that contributed most to a player's youth development.

USMNT Formative Location Dataset Validation

Before analysis, the USMNT Formative Location Dataset was validated programmatically. These checks verified that the dataset contained the six expected fields and 254 player records; contained no blank records or missing required values; contained no duplicate player names or full records; assigned each record to exactly one classification (city-coded, state-only, or International), consistent with its populated fields; and matched the documented totals of 215 city-coded, 1 state-only, and 38 International records. All programmatic checks passed.

Separately, the player-level records underwent a row-by-row consistency review against the formative-location coding procedure. This review examined proxy identification, qualifying-transition language, International classification wording, citation and linked-source alignment, and confirmation that each player's coding rationale supported the classification and location assigned to them.

Geographic Assignment and Population Baseline

Players' formative locations, recorded as city and state, were matched to their corresponding county using Geocorr 2022, a geographic correspondence engine based on 2020 Census geographies (Missouri Census Data Center 2022). For places spanning more than one county, the county with the largest population share was assigned. Geocorr matched 202 of the 215 player records automatically. The remaining 13 records, representing 12 distinct localities, were assigned manually to counties. For each locality, the county was identified using the Geographic Names Information System's Domestic Names Search Tool (U.S. Geological Survey 2022), and assigned its five-digit Federal Information Processing Standards (FIPS) code (Federal Communications Commission n.d.). Each assignment, with its supporting sources, is documented in an accompanying Manual County Resolution file.

Counties were then assigned to commuting zones (CZs), the geographic unit of analysis, using the U.S. Department of Agriculture Economic Research Service's 2020 delineation (USDA ERS 2026). After excluding Puerto Rico, this delineation contains 588 CZs, each a contiguous cluster of counties, across the 50 states and the District of Columbia. Puerto Rico was excluded because it fields a separate national team and no player in the sample had a formative location there.

This unit is granular enough to capture within-state variation but coarse enough to avoid spreading the 215 players across tens of thousands of cities. CZs also provide complete national coverage, assigning every county to a zone. Metropolitan and micropolitan delineations, by contrast, exclude counties without a strong urban core, which account for roughly 61% of all U.S. counties nationally (USDA ERS 2026).

For each CZ, the male population aged 5–19 was used as the population baseline. County-level counts were retrieved through the U.S. Census Bureau's API from the 2024 American Community Survey 5-year estimates (U.S. Census Bureau 2024). Male counts for ages 5–9, 10–14, 15–17, and 18–19 from table B01001 were summed within each county and then aggregated to the CZ level using county FIPS codes.

Statistical Analysis

Where z denotes a CZ, Observed Player Count was defined as the number of players in the analytic sample whose formative location was assigned to that zone:

$$\text{Observed Player Count}(z) = \#\{\text{players whose formative CZ} = z\}$$

Player production across CZs was summarized using the analytic sample size, the total number of CZs, the number of zones with an Observed Player Count of zero and the number with at least two, the number with at least one, and the range, mean, and median of Observed Player Count across zones. Geographic concentration was assessed across cumulative groups of CZs defined by Observed Player Count thresholds of 8, 6, 4, 2, and 1. For each group, its share of the 215 analytic-sample players was compared with its share of the nationwide male population aged 5–19.

The Population-Expected Count represents the number of players a CZ would be expected to produce if the 215 players were distributed in proportion to the male population aged 5–19. Across the 588 included CZs, this population totaled 32,532,005. Letting P(z) denote the male population aged 5–19 in CZ z, the Population-Expected Count and Observed-to-Expected Ratio for each CZ were calculated as:

$$\text{Population-Expected Count}(z) = 215 \times (P(z) \div 32{,}532{,}005)$$

$$\text{Observed-to-Expected Ratio}(z) = \text{Observed Player Count}(z) \div \text{Population-Expected Count}(z)$$

Observed-to-Expected Ratio values above 1 indicate that a CZ produced more players than expected under the population-proportional baseline, while values below 1 indicate fewer players than expected under the same baseline.

Ranking Construction

Two rankings were constructed. For the above-baseline production ranking, CZs were eligible only if Observed-to-Expected Ratio > 1 and Observed Player Count ≥ 2. The latter threshold excludes extreme Observed-to-Expected Ratio values created when a single observed player, whose presence may be incidental, is divided by a very small expected count. Eligible CZs were ranked by Observed-to-Expected Ratio, and the ten highest values were reported.

For the below-baseline production ranking, CZs were eligible only if Observed-to-Expected Ratio < 1 and Population-Expected Count ≥ 1. The Population-Expected Count threshold excludes zones expected to produce less than one player, where producing zero players is unlikely to represent a substantial shortfall. Because all zero-player zones have an Observed-to-Expected Ratio of 0 regardless of their expected counts, eligible CZs were ranked by absolute shortfall (Population-Expected Count minus Observed Player Count), and the ten largest shortfalls were reported.

Data and Code Availability

The USMNT Formative Location Dataset, the Manual County Resolution file, the validation notebook, and the analysis notebook needed to reproduce the reported results are archived in the project's Open Science Framework repository at https://osf.io/7v4w6/overview?view_only=95cc8cc414f24d0094d6b107c12a3ee9. The notebooks include the executed outputs corresponding to the tables and figure reported below.

Results

Player production was highly concentrated across CZs. Of the 588 CZs, 518 produced no players, and the median zone produced none (Table 1); the entire analytic sample originated from 70 zones.

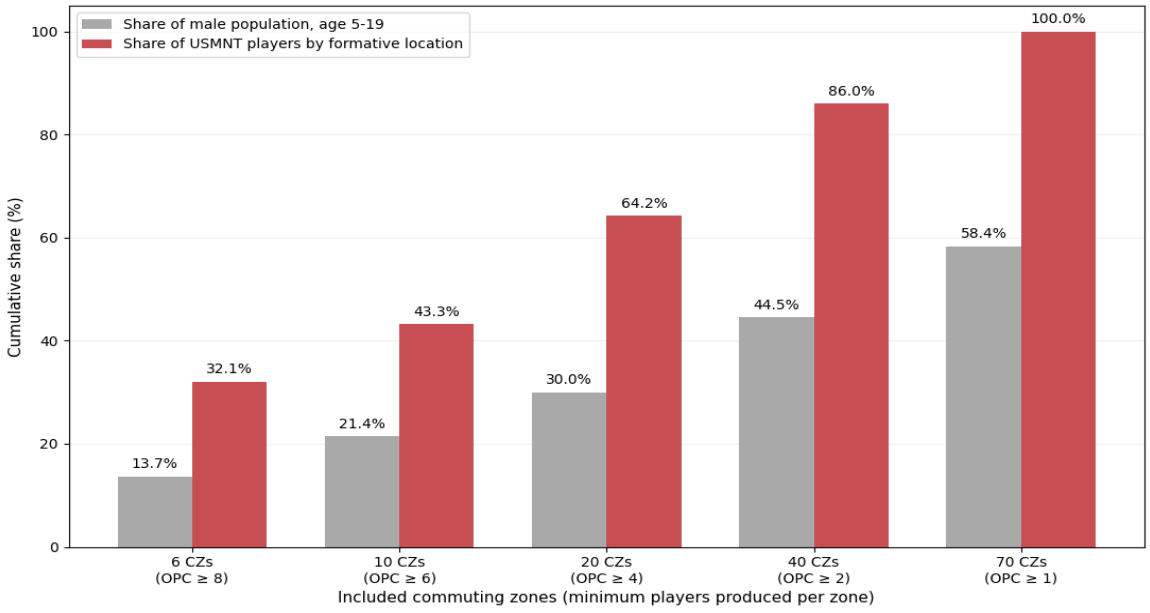
Table 1. Sample and Commuting Zone Summary Statistics

Statistic	Value
Total players (analytic sample)	215
Total CZs	588
CZs with OPC of at least 2 players	40
CZs with OPC of at least 1 player	70
CZs with OPC of 0 players	518
Range of CZ OPC	0–28
Mean CZ OPC	0.366
Median CZ OPC	0

Note: OPC denotes Observed Player Count.

Player share exceeded youth-population share at every level of aggregation: the six most productive zones accounted for 32.1% of players while containing 13.7% of the male youth population aged 5–19, and the forty zones producing at least two players accounted for 86.0% of players while containing 44.5% of that population (Figure 1).

Figure 1. Concentration of USMNT Player Production and Male Youth Population Across Successively Broader Groups of Productive Commuting Zones



Note (Figure 1): OPC denotes Observed Player Count. Groups are cumulative: each includes every commuting zone producing at least the indicated number of players, so the 10-zone group contains the 6-zone group, and so on.

The above-baseline ranking identified zones producing well above expectation (Table 2). El Paso, Germantown, and San Diego CZs recorded the highest Observed-to-Expected Ratios, each near or above four.

Table 2. Above-Baseline Production: Observed-to-Expected Ratio > 1 and Observed Player Count ≥ 2, Ranked by Observed-to-Expected Ratio

CZ	Population-Expected Count	Observed Player Count	Observed-to-Expected Ratio
El Paso, TX–NM	0.829	4	4.825
Germantown CDP, MD	1.097	5	4.556
San Diego, CA	2.027	8	3.948
Birmingham, AL	0.838	3	3.578
Honolulu, HI	0.583	2	3.433
Columbus, OH	1.472	5	3.397
St. Louis City, MO	1.558	5	3.210
Syracuse, NY	0.645	2	3.103
Washington City, DC	2.608	8	3.067
Colorado Springs, CO	0.666	2	3.002

Note: Each CZ is a group of contiguous counties. Although labels use city or place names, each row represents the entire CZ, not only the named locality. CDP denotes census-designated place.

The below-baseline ranking identified major metropolitan zones producing well below their population-based expectation (Table 3). Houston CZ produced one player against an expected 5.597, and the Boston CZ one against an expected 3.249, while San Antonio, Charlotte, Sacramento, Austin, Cincinnati, and Pittsburgh CZs each produced none despite expected counts above one.

Table 3. Below-Baseline Production: Observed-to-Expected Ratio < 1 and Population-Expected Count ≥ 1, Ranked by Population-Expected Count minus Observed Player Count

CZ	Population-Expected Count	Observed Player Count	Observed-to-Expected Ratio
Houston, TX	5.597	1	0.179
Boston, MA–NH	3.249	1	0.308
San Antonio, TX	2.158	0	0.000
Charlotte, NC–SC	1.937	0	0.000
Sacramento, CA	1.676	0	0.000
Austin, TX	1.591	0	0.000
Cincinnati, OH–KY	1.485	0	0.000
Detroit, MI	3.375	2	0.593
Pittsburgh, PA	1.327	0	0.000
Salt Lake City, UT	2.247	1	0.445

Note: Each CZ is a group of contiguous counties. Although labels use city or place names, each row represents the entire CZ, not only the named locality.

Discussion

The findings show pronounced geographic concentration in USMNT player production. Of the 588 CZs, 518 produced no players. Player share exceeded youth-population share at every threshold examined: the six most productive zones accounted for 32.1% of players while containing 13.7% of the male youth population, and the forty zones producing at least two players accounted for 86.0% of players while containing 44.5% of that population. Some CZs produced substantially more players than expected from their youth populations, while others produced substantially fewer. The El Paso, Germantown, and San Diego CZs had the highest above-baseline production, while the Houston and Boston CZs had the largest below-baseline shortfalls.

The concentration documented here makes access to development pathways an important question for subsequent research. Income-based disparities in youth sport participation are well documented (Solomon 2025). U.S. Soccer's leadership and policymakers have recently characterized youth soccer's development system as fragmented and cost-intensive, with participation that varies across parts of the country (Stynes and Tennery 2026). Therefore, a zone with a low Observed Player Count may reflect a

general scarcity of talent, or it may reflect talent that goes unidentified and undeveloped. The results of this study highlight where this distinction warrants closer investigation.

Limitations

Several limitations qualify these findings. First, formative locations were coded from publicly available biographical sources that varied in detail, and proxies were used when no direct city was documented. Second, the population baseline has its own limits: the male population aged 5–19 provides a consistent denominator but does not directly measure youth soccer participation, access, or interest, and the use of 2024 estimates does not capture population changes across the full 2010–2026 study period. Third, commuting zones can span areas with meaningfully different local conditions, and zone-level counts cannot capture that internal variation. Fourth, a zone's Observed Player Count approximates where players developed, but appearing for the USMNT also depends on later factors such as coach selection and eligibility rules; these rankings describe geographic patterns of production, not the determinants of national-team selection. Finally, because many zones produced very few players, a single player can substantially shift a zone's Observed-to-Expected Ratio. The eligibility thresholds used in the rankings reduce the most extreme cases but do not eliminate this uncertainty. Accordingly, these findings should be interpreted in context.

Conclusion

This study set out to identify the formative locations of players who appeared for the USMNT between 2010 and 2026, and to compare that geographic distribution with the nationwide distribution of the male youth population. Using a rules-based coding procedure to assign eligible players a U.S. formative location, the analysis documented a pronounced geographic imbalance: most CZs produced no players at all, a small number accounted for a share of players several times their share of the male youth population, and several CZs containing large metropolitan areas produced far fewer players than their populations would predict.

The contribution of this analysis is descriptive rather than explanatory. By covering a contemporary national-team period, benchmarking output against the male population aged 5–19, assigning every county to a zone, and ranking underproduction by absolute shortfall rather than a ratio that cannot distinguish among zero-producing zones, the study provides a systematic and reproducible account of a pattern that earlier population-relative analyses could establish only in part. The rankings identify CZs of unusually high and low production but do not establish why, and the uncertainty introduced by small player counts and imperfect biographical documentation should temper interpretation of any individual zone's position.

References

Federal Communications Commission. n.d. "Federal Information Processing System (FIPS) Codes for States and Counties." Accessed July 22, 2026. https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt.

Hill, Jamie. 2019. "The Geography of American Player Development: A Look at Where Talent Originates." *American Soccer Now*, February 21, 2019. http://americansoccernow.com/articles/the-geography-of-american-player-development-a-look-at-where-talent-originates.

Jackman, Spencer. 2017. "8 Regions Account for More Than Half of the USMNT Player Pool." *The18*, July 6, 2017. https://the18.com/en/soccer-news/8-regions-account-more-half-usmnt-player-pool/23601.

Missouri Census Data Center. 2022. "Geocorr 2022: Geographic Correspondence Engine." Version 1.8, revised October 2022. https://mcdc.missouri.edu/applications/geocorr2022.html.

Morganti, Gabriele, Adam L. Kelly, Gennaro Apollaro, Laura Pantanella, Mario Esposito, Alberto Grossi, and Bruno Ruscello. 2023. "All Roads Lead to Rome? Exploring Birthplace Effects and the 'Southern Question' in Italian Soccer." *Soccer & Society* 25 (1): 62–75. https://doi.org/10.1080/14660970.2023.2226077.

Solomon, Jon. 2025. "Participation Trends." *State of Play 2025*. Aspen Institute Project Play. https://projectplay.org/state-of-play-2025/participation-trends.

Stynes, Trevor, and Amy Tennery. 2026. "'Crisis' in Youth Sports Comes into Focus after Early US World Cup Exit." *Reuters*, July 16, 2026. https://www.reuters.com/sports/soccer/crisis-youth-sports-comes-into-focus-after-early-us-world-cup-exit-2026-07-16/.

U.S. Census Bureau. 2024. "ACS 5-Year Detailed Tables." Census Data API. Accessed July 22, 2026. https://api.census.gov/data/2024/acs/acs5.

U.S. Department of Agriculture, Economic Research Service. 2026. *2020 Commuting Zones*. January. CSV file and user's guide. https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.

U.S. Geological Survey. 2022. "Geographic Names Information System (GNIS)." November 4. https://www.usgs.gov/tools/geographic-names-information-system-gnis.

U.S. Soccer. n.d.-a. “2021 USMNT Lineups.” Accessed July 3, 2026. https://www.ussoccer.com/usmnt-lineups/2021.

U.S. Soccer. n.d.-b. “USMNT Stats.” Accessed July 3, 2026. https://www.ussoccer.com/usmnt-stats.

van Nieuwstadt, Michiel H. H., Marjolijn Das, and Marije T. Elferink-Gemser. 2021. “Mechanisms Explaining the Birthplace Effect for Male Elite Football Players.” *Journal of Sports Sciences* 39 (5): 576–82. https://doi.org/10.1080/02640414.2020.1835237.