# Effects of Playing High School Football on Mental Health in Early Adulthood: An Observational Study

Joint work with Sameer Deshpande*, Jordan Weiss, and Dylan Small**

*First Author; ** Principal Investigator

Credit: flisk Getty Images

Wharton
University of Pennsylvania

Add Health
The National Longitudinal Study of Adolescent to Adult Health

# Epidemiological importance of studying the health effects of HS football

# Epidemiological importance of studying the health effects of HS football

- **Impact:** Largest participation sport in US high schools — over 1 million boys played HS football in 2014.

# Epidemiological importance of studying the health effects of HS football

- **Impact:** Largest participation sport in US high schools — over 1 million boys played HS football in 2014.

- **Known risks:** Leading cause of concussions among adolescents and young adults — account for nearly half of all sports-related concussions (Marar et al., 2012).

# Epidemiological importance of studying the health effects of HS football

- **Impact:** Largest participation sport in US high schools — over 1 million boys played HS football in 2014.

- **Known risks:** Leading cause of concussions among adolescents and young adults — account for nearly half of all sports-related concussions (Marar et al., 2012).

- **Increasing public awareness:** High profile studies and news coverage linking Chronic Traumatic Encephalopathy (CTE) and football has brought the issue of safety in HS football to the fore.

# Epidemiological importance of studying the health effects of HS football

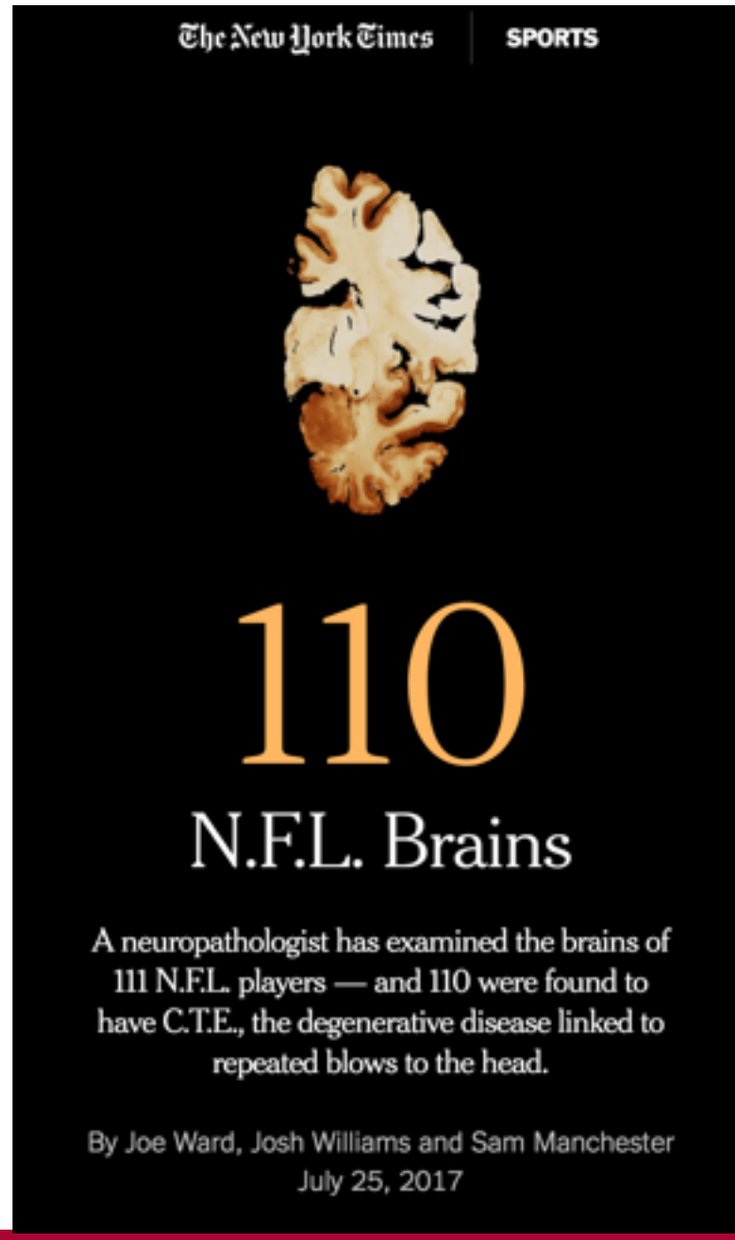
- **Impact:** Largest participation sport in US high schools — over 1 million boys played HS football in 2014.

- **Known risks:** Leading cause of concussions among adolescents and young adults — account for nearly half of all sports-related concussions (Marar et al., 2012).

- **Increasing public awareness:** High profile studies and news coverage linking Chronic Traumatic Encephalopathy (CTE) and football has brought the issue of safety in HS football to the fore.

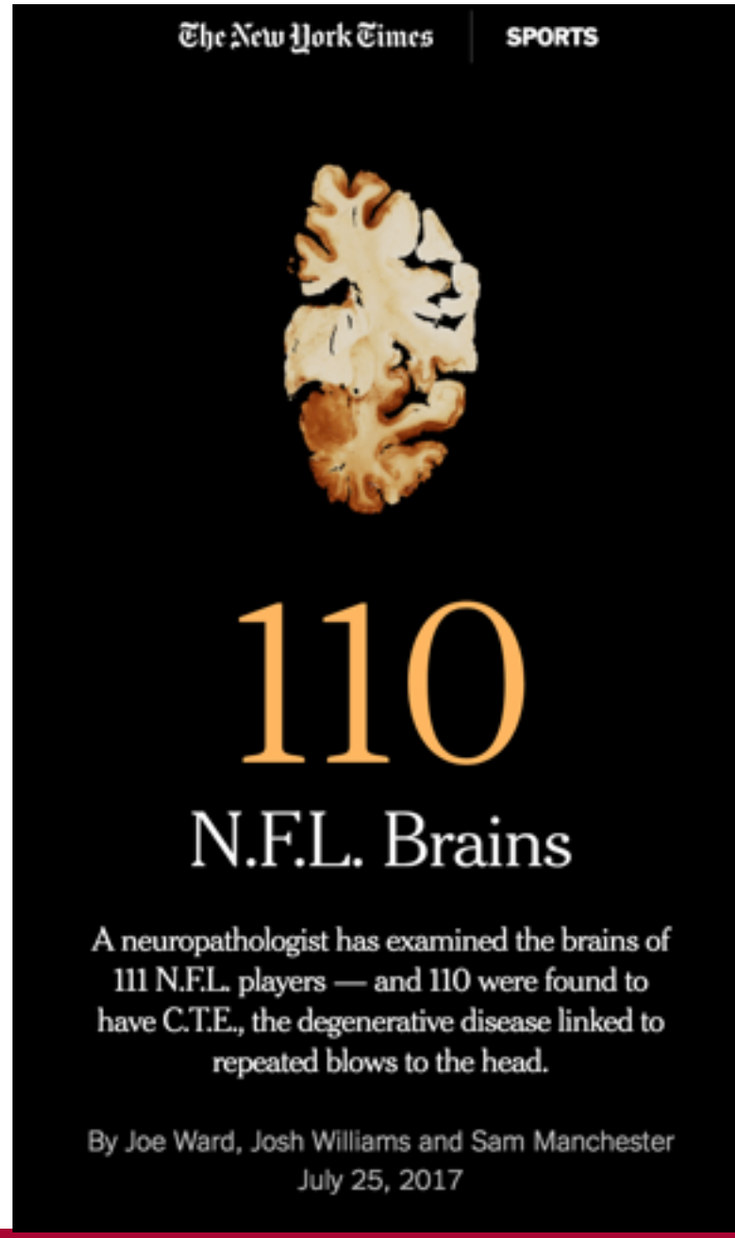
- **Current evidence:** remains inconclusive…

# Current state of the evidence: Post-mortem CTE studies



# Current state of the evidence: Post-mortem CTE studies

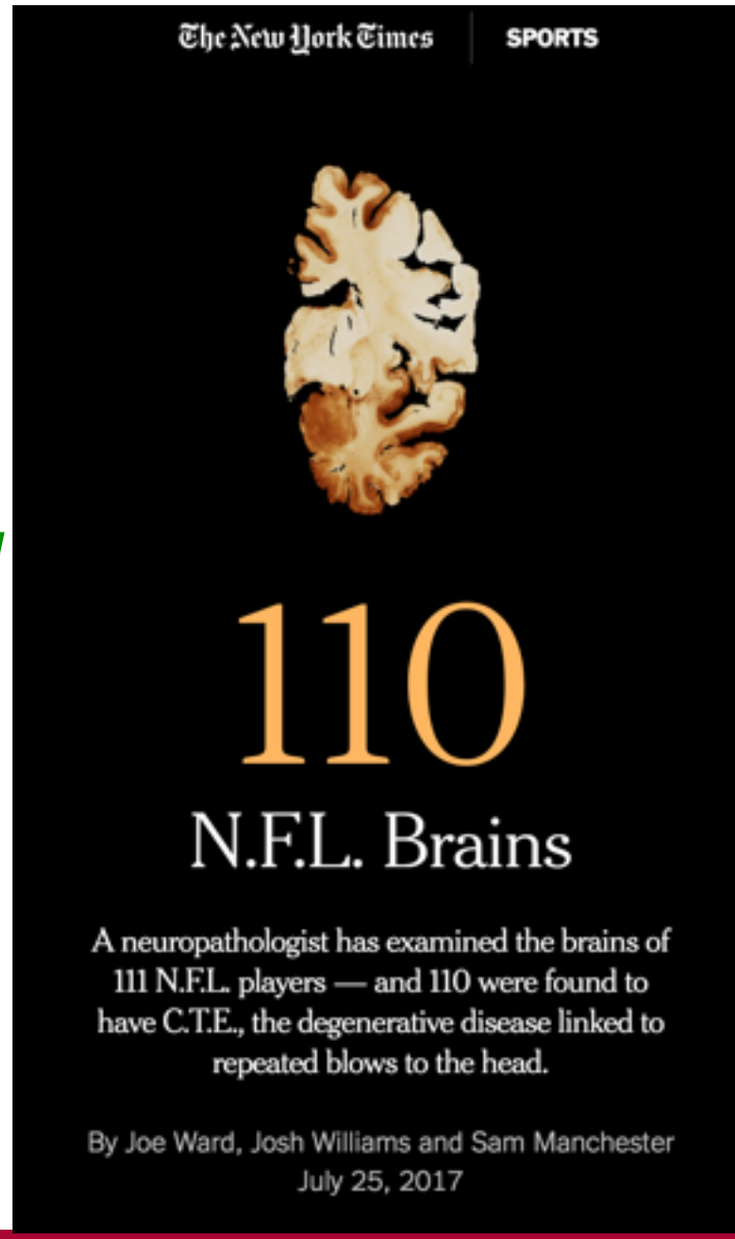
- **Boston University Brain Bank Study (BU): "**CTE was neuropathologically diagnosed in 177 players across all levels of play [among a convenience sample of 202 deceased players of American football] from a brain donation program, including 110 of 111 NFL players," (McKee et al., 2017).



# Current state of the evidence: Post-mortem CTE studies

- **Boston University Brain Bank Study (BU): "**CTE was neuropathologically diagnosed in 177 players across all levels of play [among a convenience sample of 202 deceased players of American football] from a brain donation program, including 110 of 111 NFL players," (McKee et al., 2017).

  + *(1)* ***Accurate measure of exposure and dose (level and years of play, history of head trauma)****, (2) considers a neurophysiological pathway for clinical symptoms.*

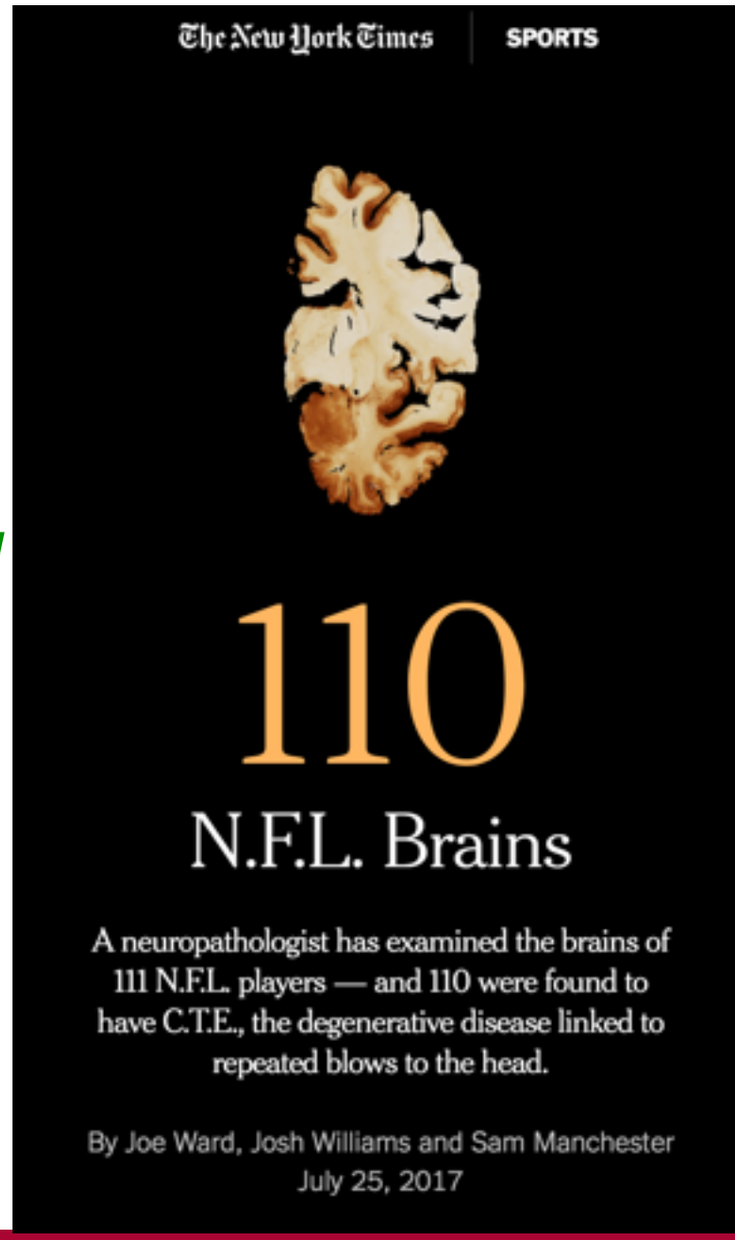


# Current state of the evidence: Post-mortem CTE studies

- **Boston University Brain Bank Study (BU): "**CTE was neuropathologically diagnosed in 177 players across all levels of play [among a convenience sample of 202 deceased players of American football] from a brain donation program, including 110 of 111 NFL players," (McKee et al., 2017).

  + *(1)* ***Accurate measure of exposure and dose (level and years of play, history of head trauma)****, (2) considers a neurophysiological pathway for clinical symptoms.*

  - *(1)* ***Severe referral bias — donation to brain bank associated with dementia and depression status****, (2) neuropathology overlaps broadly with several neurodegenerative comorbidities (Schwab and Hazrati, 2018), (3) Does not generalize well to most common levels of exposure (i.e. youth and HS football).*



# Current state of the evidence: Longitudinal epidemiological studies

ROCHESTER EPIDEMIOLOGY PROJECT
EST. 1966
A Minnesota and Wisconsin Collaboration

WISCONSIN LONGITUDINAL STUDY

Wharton
University of Pennsylvania

Add Health
The National Longitudinal Study of Adolescent to Adult Health

# Current state of the evidence: Longitudinal epidemiological studies

- **Rochester Epidemiology Project (REP):** 438 male students who played HS football between 1946-1956 found to have **no significant increased incidence** of dementia, PD, or ALS when compared to 140 non-football playing male classmates or general population incidence rates (Savica et al., 2012).

ROCHESTER EPIDEMIOLOGY PROJECT
EST. 1966
A Minnesota and Wisconsin Collaboration

WISCONSIN LONGITUDINAL STUDY

Wharton
University of Pennsylvania

Add Health
The National Longitudinal Study of Adolescent to Adult Health

# Current state of the evidence: Longitudinal epidemiological studies

- **Rochester Epidemiology Project (REP):** 438 male students who played HS football between 1946-1956 found to have **no significant increased incidence** of dementia, PD, or ALS when compared to 140 non-football playing male classmates or general population incidence rates (Savica et al., 2012).

  + *(1) Representative sample, (2)* ***linked to medical records and diagnoses of neurodegenerative disease.***

ROCHESTER EPIDEMIOLOGY PROJECT EST. 1966
A Minnesota and Wisconsin Collaboration

WISCONSIN LONGITUDINAL STUDY

Wharton University of Pennsylvania

Add Health The National Longitudinal Study of Adolescent to Adult Health

# Current state of the evidence: Longitudinal epidemiological studies

- **Rochester Epidemiology Project (REP):** 438 male students who played HS football between 1946-1956 found to have **no significant increased incidence** of dementia, PD, or ALS when compared to 140 non-football playing male classmates or general population incidence rates (Savica et al., 2012).

  \+ *(1) Representative sample, (2)* ***linked to medical records and diagnoses of neurodegenerative disease.***

  \- *(1)* ***No adjustment for important confounders*** *(e.g. adolescent IQ, general health, etc.), (2) relevance of exposure to current players, (3) no measure of dose.*

ROCHESTER EPIDEMIOLOGY PROJECT EST. 1966
A Minnesota and Wisconsin Collaboration

WISCONSIN LONGITUDINAL STUDY

Wharton University of Pennsylvania

Add Health The National Longitudinal Study of Adolescent to Adult Health

# Current state of the evidence: Longitudinal epidemiological studies






# Current state of the evidence: Longitudinal epidemiological studies

- **Wisconsin Longitudinal Study (WLS):** From random sample of 1957 HS graduates in WI, **no significant harmful effect** on later-life cognition and depression detected for 834 football players compared to 566 non-collision athletes and 1292 non-athletes (Deshpande et al., 2017).






# Current state of the evidence: Longitudinal epidemiological studies

- **Wisconsin Longitudinal Study (WLS):** From random sample of 1957 HS graduates in WI, **no significant harmful effect** on later-life cognition and depression detected for 834 football players compared to 566 non-collision athletes and 1292 non-athletes (Deshpande et al., 2017).

  + *(1) Representative sample, (2) **controlled for rich set of baseline covariates** (e.g. adolescent IQ, APOE status, etc.), (3) multiple control groups designed to address certain sources of unmeasured confounding*




# Current state of the evidence: Longitudinal epidemiological studies

- **Wisconsin Longitudinal Study (WLS):** From random sample of 1957 HS graduates in WI, **no significant harmful effect** on later-life cognition and depression detected for 834 football players compared to 566 non-collision athletes and 1292 non-athletes (Deshpande et al., 2017).

+ *(1) Representative sample, (2) **controlled for rich set of baseline covariates** (e.g. adolescent IQ, APOE status, etc.), (3) multiple control groups designed to address certain sources of unmeasured confounding*

- *(1) Relevance of exposure to current players, (2) **Sensitivity of cognitive outcome measures to mTBI-related disease**, (3) no measure of dose.*

# Why Add Health?

# Why Add Health?

- **Implications that generalize:** Nationally representative population sample of adolescents in grades 7-12 in 1994-95.

# Why Add Health?

- **Implications that generalize:** Nationally representative population sample of adolescents in grades 7-12 in 1994-95.

- **Relevant Exposure:** Recorded whether student played or intended to play HS football and playing HS football in 1990s better approximates exposure for current and future players than WLS study.

# Why Add Health?

- **Implications that generalize:** Nationally representative population sample of adolescents in grades 7-12 in 1994-95.

- **Relevant Exposure:** Recorded whether student played or intended to play HS football and playing HS football in 1990s better approximates exposure for current and future players than WLS study.

- **Possible effects in early adulthood:** Several case studies found CTE in college players who committed suicide (Schwarz, 2010; Bever, 2014; Bishop, 2018) —> Focus on Wave IV mental health outcomes measured between ages 24 and 32.

# Why Add Health?

- **Implications that generalize:** Nationally representative population sample of adolescents in grades 7-12 in 1994-95.

- **Relevant Exposure:** Recorded whether student played or intended to play HS football and playing HS football in 1990s better approximates exposure for current and future players than WLS study.

- **Possible effects in early adulthood:** Several case studies found CTE in college players who committed suicide (Schwarz, 2010; Bever, 2014; Bishop, 2018) —> Focus on Wave IV mental health outcomes measured between ages 24 and 32.

- **Adjusting for important confounders:** availability of rich set of important baseline covariates.

# Constructing the study sample and defining the exposure

# Constructing the study sample and defining the exposure

- **Sample:** 5,780 men in core Add Health sample, 2,877 answered in-school questionnaire in Wave I.

# Constructing the study sample and defining the exposure

- **Sample:** 5,780 men in core Add Health sample, 2,877 answered in-school questionnaire in Wave I.

- **Exclusions:** 993 men indicated participation or intent to participate in another sport associated with high risk of repeated head trauma (e.g. soccer or wrestling); 119 deemed to have a physical disability or limitation.

# Constructing the study sample and defining the exposure

- **Sample:** 5,780 men in core Add Health sample, 2,877 answered in-school questionnaire in Wave I.

- **Exclusions:** 993 men indicated participation or intent to participate in another sport associated with high risk of repeated head trauma (e.g. soccer or wrestling); 119 deemed to have a physical disability or limitation.

- **Exposure:** indicated in Wave I questionnaire that they participated or intended to participate in HS football (n=661 men in core Add Health sample).

# Constructing the study sample and defining the exposure

- **Sample:** 5,780 men in core Add Health sample, 2,877 answered in-school questionnaire in Wave I.

- **Exclusions:** 993 men indicated participation or intent to participate in another sport associated with high risk of repeated head trauma (e.g. soccer or wrestling); 119 deemed to have a physical disability or limitation.

- **Exposure:** indicated in Wave I questionnaire that they participated or intended to participate in HS football (n=661 men in core Add Health sample).

- **Control:** 2,216 control subjects, 808 participated or intended to participate in a non-collision sport (e.g. basketball or swimming) and 1,408 were non-athletes.

# Mental health consequences of HS football: Primary and secondary outcomes

# Mental health consequences of HS football: Primary and secondary outcomes

- **Primary outcome:** Center for Epidemiological Studies' Depression Scale (CES-D) score (Radloff, 1977) measured in Wave IV.

# Mental health consequences of HS football: Primary and secondary outcomes

- **Primary outcome:** Center for Epidemiological Studies' Depression Scale (CES-D) score (Radloff, 1977) measured in Wave IV.

- **Secondary outcomes:**

# Mental health consequences of HS football: Primary and secondary outcomes

- **Primary outcome:** Center for Epidemiological Studies' Depression Scale (CES-D) score (Radloff, 1977) measured in Wave IV.

- **Secondary outcomes:**

  ★ *General health (e.g. diagnosed with heart disease after age 18)*

# Mental health consequences of HS football: Primary and secondary outcomes

- **Primary outcome:** Center for Epidemiological Studies' Depression Scale (CES-D) score (Radloff, 1977) measured in Wave IV.

- **Secondary outcomes:**

  ★ *General health (e.g. diagnosed with heart disease after age 18)*

  ★ *Substance dependence/abuse (e.g. lifetime diagnosis of alcohol dependence/abuse)*

# Mental health consequences of HS football: Primary and secondary outcomes

- **Primary outcome:** Center for Epidemiological Studies' Depression Scale (CES-D) score (Radloff, 1977) measured in Wave IV.

- **Secondary outcomes:**

  ★ *General health (e.g. diagnosed with heart disease after age 18)*

  ★ *Substance dependence/abuse (e.g. lifetime diagnosis of alcohol dependence/abuse)*

  ★ *Personality (e.g. Angry Hostility Scale)*

# In lieu of experiment: Good observational design

# In lieu of experiment: Good observational design

- **Randomization as "reasoned basis for inference" (Fisher):** Construct matched sets that are relatively homogenous in propensity score and baseline covariates to approximate a stratified experiment.

# In lieu of experiment: Good observational design

- **Randomization as "reasoned basis for inference" (Fisher):** Construct matched sets that are relatively homogenous in propensity score and baseline covariates to approximate a stratified experiment.

- **Anticipate unmeasured confounding:** Consider multiple comparisons that address plausible sources of unmeasured confounding while not sacrificing power.

# In lieu of experiment: Good observational design

- **Randomization as "reasoned basis for inference" (Fisher):** Construct matched sets that are relatively homogenous in propensity score and baseline covariates to approximate a stratified experiment.

- **Anticipate unmeasured confounding:** Consider multiple comparisons that address plausible sources of unmeasured confounding while not sacrificing power.

- **Avoid data-snooping:** Like a RCT, publicly register a pre-analysis plan prior to joint analysis of outcomes and exposure (Rubin, 2007).

# Matching on important baseline covariates and estimated propensity scores [1/2]

# Matching on important baseline covariates and estimated propensity scores [1/2]

- **Approximating a stratified experiment:** Create strata by dividing into grade groups (7-8, 9-10, and 11-12) then matching treated subjects closely on important baseline covariates and propensity scores to a variable number of controls.

# Matching on important baseline covariates and estimated propensity scores [1/2]

- **Approximating a stratified experiment:** Create strata by dividing into grade groups (7-8, 9-10, and 11-12) then matching treated subjects closely on important baseline covariates and propensity scores to a variable number of controls.

★ *Conditioning on propensity scores* ***balances covariate distributions*** *between treated and controls (Rosenbaum and Rubin, 1983).*

# Matching on important baseline covariates and estimated propensity scores [1/2]

- **Approximating a stratified experiment:** Create strata by dividing into grade groups (7-8, 9-10, and 11-12) then matching treated subjects closely on important baseline covariates and propensity scores to a variable number of controls.

★ *Conditioning on propensity scores* ***balances covariate distributions*** *between treated and controls (Rosenbaum and Rubin, 1983).*

★ *Matching closely on covariates guards against misspecification of propensity score and* ***reduces variance****.*

# Matching on important baseline covariates and estimated propensity scores [1/2]

- **Approximating a stratified experiment:** Create strata by dividing into grade groups (7-8, 9-10, and 11-12) then matching treated subjects closely on important baseline covariates and propensity scores to a variable number of controls.

  ★ *Conditioning on propensity scores* ***balances covariate distributions*** *between treated and controls (Rosenbaum and Rubin, 1983).*

  ★ *Matching closely on covariates guards against misspecification of propensity score and* ***reduces variance****.*

- **Rich set of baseline covariates:** sociodemographic status, school performance/engagement, life plan, social engagement, psychological characteristics, physical functioning.

# Versions of control: Systematic variation of unmeasured confounders

# Versions of control: Systematic variation of unmeasured confounders

- **A natural comparison:** Football players vs. all male students that did not play football (*all controls*).

# Versions of control: Systematic variation of unmeasured confounders

- **A natural comparison:** Football players vs. all male students that did not play football (*all controls*).

- **An oversimplification?** Potentially two *versions* of control: controls who played non-collision sports and those who played no sports at all — may differ in important, but unmeasured dimensions like personality, temperament, and fitness. Comparing separately to both groups systematically varies these potential confounders.

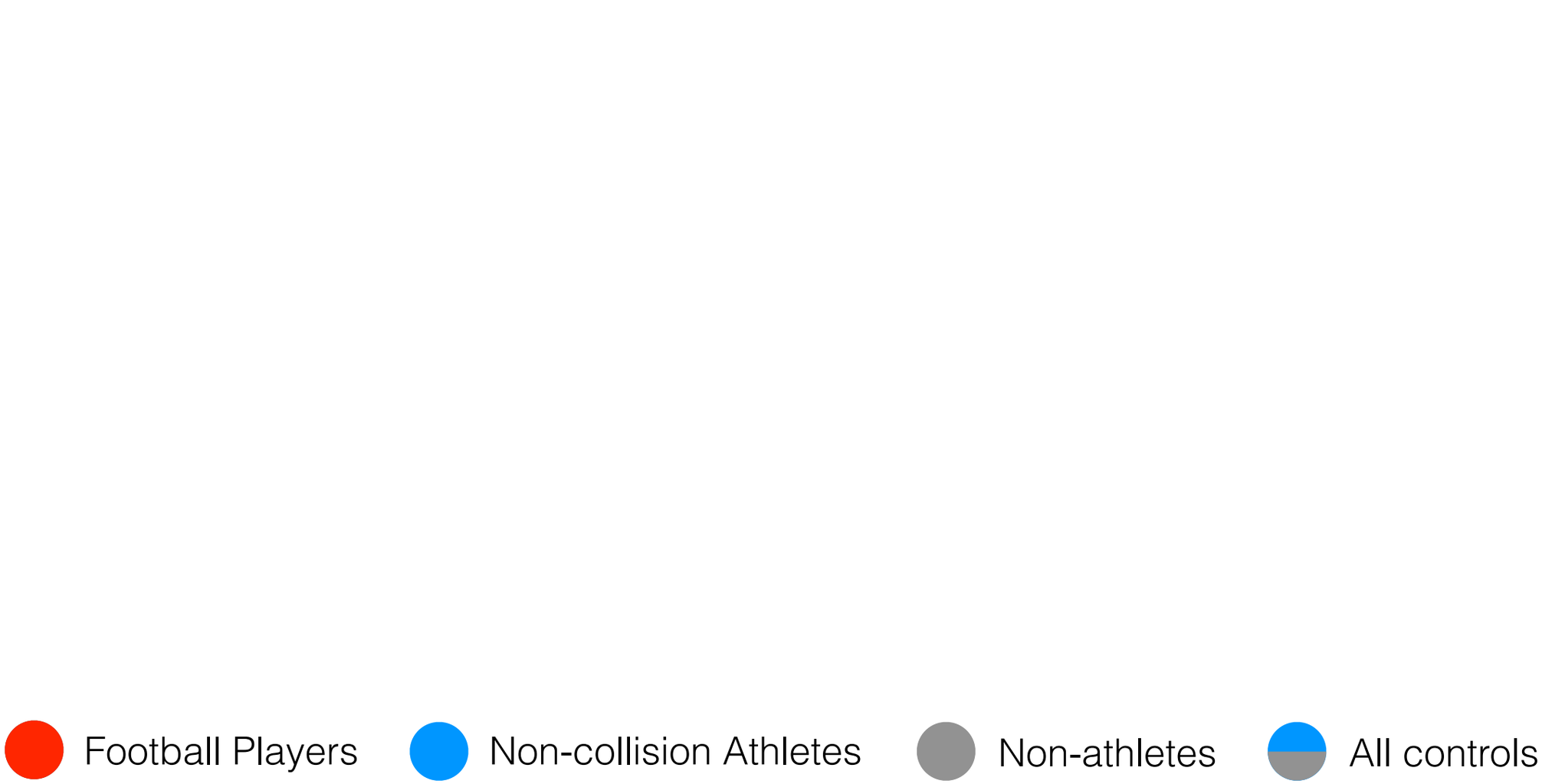
# Versions of control: Systematic variation of unmeasured confounders

- **A natural comparison:** Football players vs. all male students that did not play football (*all controls*).

- **An oversimplification?** Potentially two *versions* of control: controls who played non-collision sports and those who played no sports at all — may differ in important, but unmeasured dimensions like personality, temperament, and fitness. Comparing separately to both groups systematically varies these potential confounders.

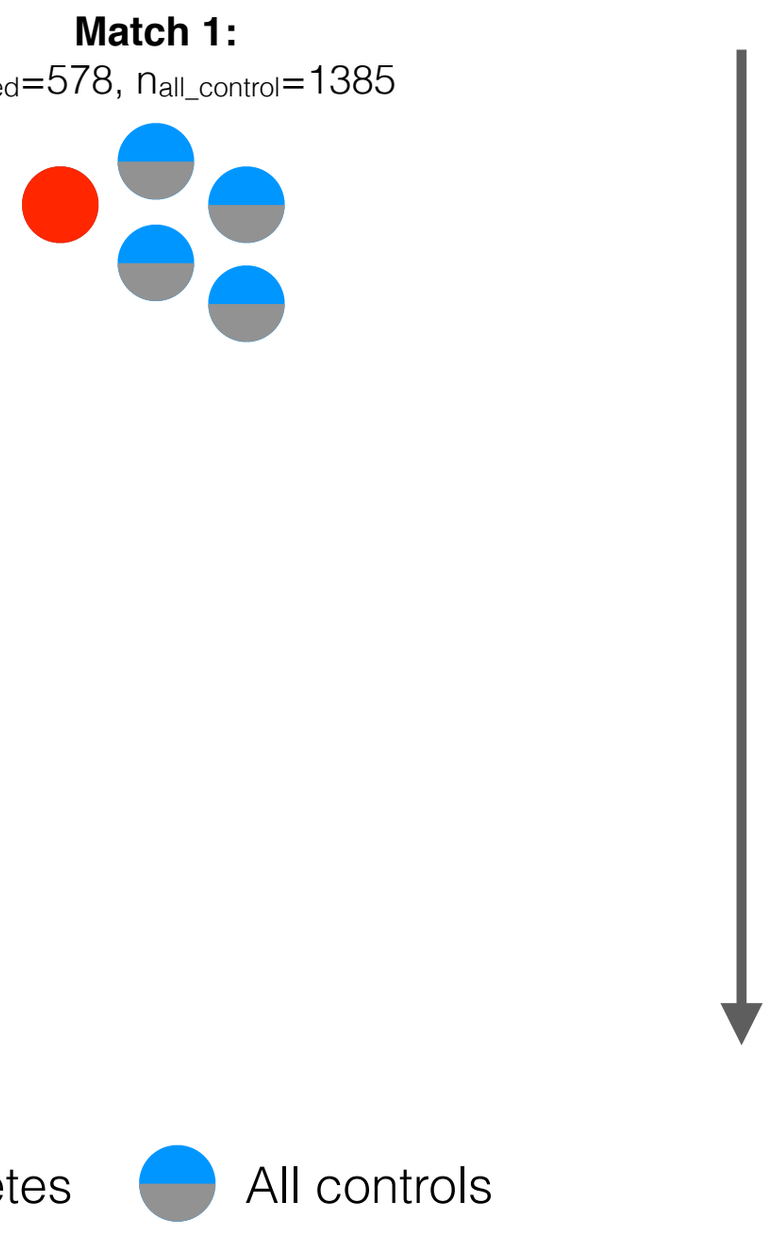
- **Solution:** Developed an ordered testing procedure that seeks reassurance that the natural comparison is not an oversimplification of these *versions* while not sacrificing power in the main comparison (Hasegawa et al., 2018).

# Testing in order: all controls, versions of control, and equivalence of controls



# Testing in order: all controls, versions of control, and equivalence of controls

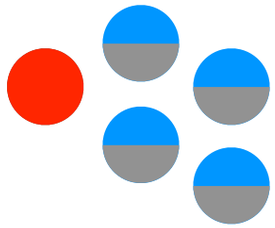
- **Match 1:** treatment vs. all controls — maximize power.



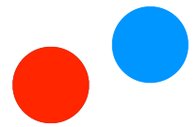
# Testing in order: all controls, versions of control, and equivalence of controls

- **Match 1:** treatment vs. all controls — maximize power.
- **Match 2 & 3:** treatment vs. non-collision athletes & treatment vs. non-athletes — are athletes inherently different than non-athletes? do positive effects of playing a sport counteract negative effects from head-injury exposure?

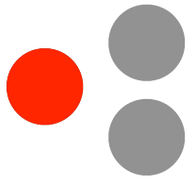
**Match 1:**
$n_{treated}$=578, $n_{all\_control}$=1385



**Match 2:**
$n_{treated}$=444, $n_{non\text{-}collision}$=606



**Match 3:**
$n_{treated}$=411, $n_{non\text{-}athlete}$=788



Football Players  Non-collision Athletes  Non-athletes  All controls

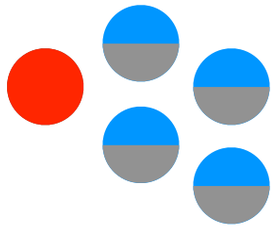
# Testing in order: all controls, versions of control, and equivalence of controls

- **Match 1:** treatment vs. all controls — maximize power.

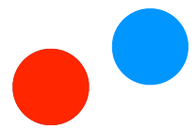
- **Match 2 & 3:** treatment vs. non-collision athletes & treatment vs. non-athletes — are athletes inherently different than non-athletes? do positive effects of playing a sport counteract negative effects from head-injury exposure?

- **Match 4:** equivalence test of non-collision athletes vs. non-athletes — seek further reassurance that the comparison in Match 1 is appropriate.

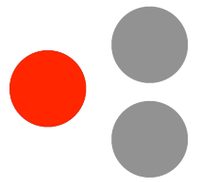
**Match 1:**
$n_{treated}$=578, $n_{all\_control}$=1385



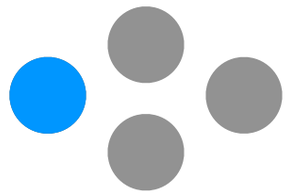
**Match 2:**
$n_{treated}$=444, $n_{non\text{-}collision}$=606



**Match 3:**
$n_{treated}$=411, $n_{non\text{-}athlete}$=788



**Match 4:**
$n_{non\text{-}collision}$= 596, $n_{non\text{-}athlete}$= 1047



Football Players  Non-collision Athletes  Non-athletes  All controls

# Results… Coming Soon!

# Results… Coming Soon!

- Without experimental control, good science takes even greater patience and restraint from researchers to avoid data snooping, p-hacking, etc.

# Results… Coming Soon!

- Without experimental control, good science takes even greater patience and restraint from researchers to avoid data snooping, p-hacking, etc.
- With that in mind, we are currently finalizing a pre-analysis protocol, which we will register at clinicaltrials.gov and post on arXiv.org **prior** to analyzing exposure and outcome data together.

# Results… Coming Soon!

- Without experimental control, good science takes even greater patience and restraint from researchers to avoid data snooping, p-hacking, etc.

- With that in mind, we are currently finalizing a pre-analysis protocol, which we will register at clinicaltrials.gov and post on arXiv.org **prior** to analyzing exposure and outcome data together.

- Open to advice, comments, and suggestions from experts on Add Health to help us refine our analysis plan.

# Thank you!

Contact Info:
raiden@wharton.upenn.edu
sameerd@alum.mit.edu
dsmall@wharton.upenn.edu

# BACKUP SLIDE [1/2]: An additional feature of our observational study design

- **Minimize variance without making additional modeling assumptions:** Agnostic covariance adjustment using flexible machine learning algorithms (Rosenbaum, 2002).

# BACKUP SLIDE [2/2]: Flexible covariance adjustment: Bayesian Additive Regression Trees (BART)

- Under hypothesis of an additive treatment effect $\tau$ can infer potential outcomes had all subjects received control.

- Can use machine learning algorithms to "model" potential outcomes under control to reduce variance coming from residual imbalances in prognostic covariates after matching.

- Correctness of "model" not required for valid inference, that comes from randomization or like-randomized observational study.

- BART is a flexible, Bayesian regression algorithm with nice regularization properties.