SUPPLEMENTAL MATERIALS

Who Counts in Prostate Cancer Epidemiology?: Rethinking Sex, Gender, & Risk

Developed in Women, Gender and Health 207:
Advanced Topics of Women, Gender, and Health,
Harvard T.H. Chan School of Public Health, Spring 2026

Course Instructed by Sabra L. Katz-Wise and Ann Caroline Danielsen

Teaching Example Authored by Julia Naganuma-Carreras, Ashlynn Wimer, and Kris Berg

Case Study:

A research team is using electronic health record (EHR) data from a large health system in the United States to study prostate cancer incidence and mortality. The study population is defined as "male patients ages 45+" based on the sex field in the EHR. Available variables include age, race/ethnicity, socioeconomic status, comorbidities, and treatment information, but there is no direct variable for anatomical inventory (e.g., presence of a prostate) or gender identity.

In their preliminary analysis, the team finds that Black patients have higher prostate cancer mortality than White patients, even after adjusting for age and selected clinical factors. They also notice that a small number of prostate cancer diagnoses appear among patients recorded as female, but these cases are excluded because they do not meet the study's "male patient" eligibility criteria.

The team concludes that the racial mortality disparity may reflect biological differences in disease aggressiveness and that the excluded cases are likely coding errors.

Discussion Questions:

- How did the researchers define the population at risk?
- What kinds of misclassification could be occurring?
- What are the potential health consequences of such misclassification?
- What variables would improve the study design?
- What forms of differential exposure might explain the racial mortality disparity?

Key Takeaways:

- **Cisnormativity and population at risk:** Cisnormative assumptions—defining risk as "male patients" and treating the EHR sex field as equivalent to prostate status—can (re)produce biased epidemiologic results by excluding some people with prostates (e.g., many transgender women) and including some without (e.g., transgender men).
- **Misclassification:** Using sex (or gender) as a proxy for prostate status and discarding "female" prostate cancer cases as errors, misclassifies both the population at risk and outcomes.

- **Health consequences:** Misclassification can distort incidence and mortality estimates, hide risks for transgender women, intersex, and nonbinary people, and misdirect screening, diagnosis, and treatment resources.
- **Race and biology:** Attributing higher mortality in Black patients to biology ignores social factors (e.g., structural racism and social determinants) that may drive disparities.

Glossary:

- **Anatomical inventory / anatomy-based risk:** A method of identifying which organs or tissues a person currently has, rather than assuming anatomy based on sex assigned at birth or gender identity. For prostate cancer, the key question is whether a person has a prostate. This helps define the true population at risk and avoid excluding or misclassifying transgender, nonbinary, and intersex people.
- **Cisgender:** A person whose gender identity aligns with the sex they were assigned at birth. For example, a person assigned male at birth who identifies as a man is a cisgender man. In prostate cancer epidemiology, cisgender men are usually treated as the default population of interest, but they are not the only people who may have prostates.
- **Cisnormativity:** The assumption that all dimensions of sex and gender are concordant within individuals and consistent over the life course.
- **Gender identity:** A person's internal sense of gender. It is related to, but not the same as, sex assigned at birth. When gender identity aligns with sex assigned at birth, a person may be described as cisgender; when it differs, a person may be described as transgender or nonbinary. Gender identity matters epidemiologically because it influences how people are categorized in records, how they interact with health systems, whether they receive appropriate screening, and whether they are visible in research.
- **Gender minority:** A broad academic term for people whose gender identity, expression, or experience differs from dominant societal expectations based on sex assigned at birth. This can include transgender, nonbinary, genderqueer, Two-Spirit, and other gender-diverse people.
- **Sex (assigned at birth):** The classification recorded at birth, usually female, male, or sometimes intersex (though birth certificates typically list either female or male), based on external anatomy and/or other sex characteristics including chromosomes, hormones, gonads, internal reproductive organs, and external genitalia. Sex assigned at birth is relevant to prostate cancer because people assigned male at birth are usually born with a prostate. However, sex assigned at birth should not be used as a substitute for gender identity, current anatomy, or anatomical inventory. Some intersex people may have prostates depending on their specific sex characteristics and medical history, so prostate cancer risk should be assessed based on anatomy rather than sex category alone.
- **Single-axis analysis:** An approach that examines one social category at a time, such as race/ethnicity alone or gender identity alone. Single-axis analysis can be useful, but it may miss disparities that occur at the intersection of multiple identities. For example, analyzing prostate cancer only by race/ethnicity may overlook transgender women with prostates; analyzing only by gender identity may overlook racialized differences within transgender populations.
- **SOGI(E) data:** An abbreviation for sexual orientation and gender identity (and expression) data. In this example, gender identity data are important because they help researchers distinguish between cisgender and transgender populations.
- **Transfeminine:** Transfeminine (often abbreviated as "transfem" or "transfemme") is an umbrella term for people assigned male at birth (AMAB) who have a feminine gender identity or expression, typically including transgender women and some nonbinary people.

- **Transgender:** A person whose gender identity differs from the sex they were assigned at birth. Transgender is an umbrella term and can include transgender women, transgender men, nonbinary people, or other groups depending upon the relevant social context.
- **Transmasculine:** Transmasculine (often abbreviated as "transmasc") is an umbrella term for people assigned female at birth (AFAB) who have a masculine gender identity or expression, typically including transgender men and some nonbinary people.

Who Counts in Prostate Cancer Epidemiology?: Rethinking Sex, Gender, & Risk

By Julia Naganuma-Carreras, Ashlynn Wimer, and Kris Berg

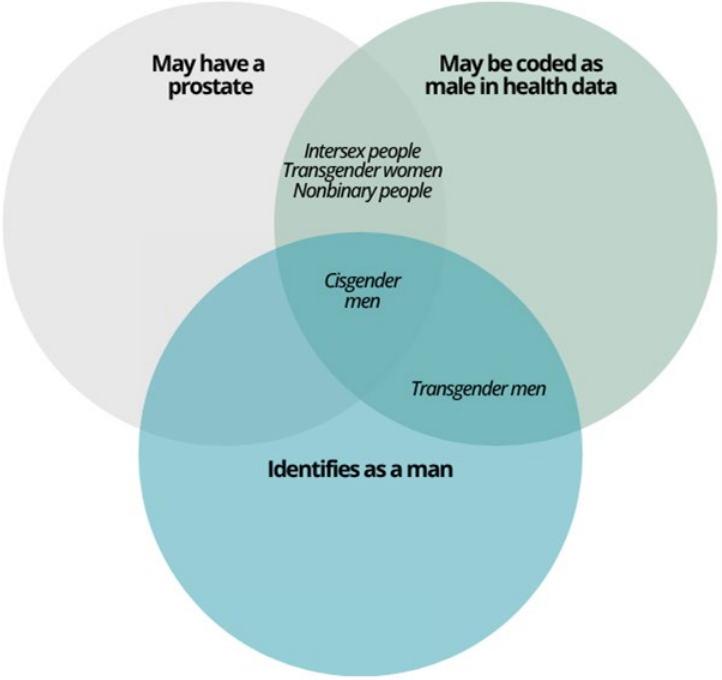
“What are these slides?”

These slides are meant to serve as templates for a course, with the goal of being hot swappable with already existent slideshows. Each slide has extensive speaker notes explaining the visual content of the slides. Feel free to use or remix as needed – everything produced here is under a Creative Commons license.

Who gets counted in cancer epidemiology?

Background

- Prostate cancer is often described as a cancer affecting “men”
- Epidemiologically, the population at risk is more precise: *peopl*
  with prostates
- “Men” is an imprecise proxy for prostate cancer risk
  - Some men, including many transmasculine people, *do not* have prostates
  - Some women, including many transfeminine people, *do* have prostates



Significance

- Incidence estimates depend on the correct denominator
- Screening depends on identifying who is actually at risk
- Studies relying only on binary sex/gender may exclude or misclassify relevant populations

Prostate cancer risk factors outside of men

- Transgender women* also have prostates and are at risk of prostate cancer
- Little is known about the epidemiology of prostate cancer among this group

Brief Report FREE

Prevalence and Factors Associated With Prostate Cancer Among Transgender Women

Celeste Manfredi, MD[1,2]; Antonio Franco, MD[1,3]; Francesco Ditonno, MD[1,4]; Eugenio Bologna, MD[1,5]; Leslie Claire Licari, MD[1,5]; Costantino Leonardo, MD[5]; Alessandro Antonelli, MD, PhD[4]; Cosimo De Nunzio, MD, PhD[3]; Edward E. Cherullo, MD, PhD[1]; Marco De Sio, MD, PhD[2]; Riccardo Autorino, MD, PhD[1]

Transgender women are people assigned male at birth and who identify as women

Table 2. Factors Associated With PCa in Transgender Women

Variable	Adjusted OR (95% CI)[a]	*P* value[b]
Age, y	1.10 (1.08-1.12)	<.001
Family history of PCa		
No	1 [Reference]	<.001
Yes	2.27 (1.60-4.92)	
Orchiectomy		
No	1 [Reference]	0.22
Yes	1.88 (0.93-3.26)	
GAHT		
No	1 [Reference]	<.001
Yes	0.60 (0.56-0.89)	

Abbreviations: CCI, Charlson Comorbidity Index; GAHT, gender-affirming hormone therapy; PCa, prostate cancer; SDOH, social determinants of health.

[a] Adjusted for age, CCI, SDOH, family history of PCa, orchiectomy, GAHT.

[b] Multivariable logistic regression analysis was used to detect the factors associated with PCa.

- Retrospective cohort of 95,460 transgender women
- **Assembled via commercially available claims dataset**
- Two "Letters & Responses" were submitted raising concerns about bias introduced from claims data

Note: GAHT means gender affirming hormone therapy, which for transgender women typically consists of endogenous estrogen.

eTable 1. Codes used to identify the required data

	ICD-9	ICD-10	CPT	USC
Transsexualism	302.50, 302.51, 302.52, 302.53, 302.6, 302.85	F64.0, F64.1, F64.2, F64.8, F64.9, Z87.890	55970	/
Orchiectomy	62.41	0VTC0ZZ, 0VTC4ZZ	54520, 54690	/
Penectomy	64.3	0VTS0ZZ, 0VTSXZZ	54125	/
Vaginoplasty	64.5	0W4M070, 0W4M0J0, 0W4M0K0	57335, 57291, 57292, 56805, 58999, 56800	/
GAHT	V07.59, V07.51, V87.43	Z79.810, Z79.818, Z92.23	/	/
PCa	185, V10.46	C61, Z85.46	/	/
Family history of PCa	Z80.42	V1642	/	/
PSA	/	/	84153	/
BCR	/	R97.21	/	/
PM	198.5	C79.51	/	/

What biases may be introduced through the usage of retrospective claims data?

Key takeaways

- Prostate cancer among transgender women highlights broader methodologic issues in cancer epidemiology.
- Study design and cohort selection can bias findings when “men” or “males” is used as a proxy for “people with prostates.”
- Better data collection should match what is being measured, including anatomy, gender identity, sex assigned at birth, hormones, and surgical history.
- Research should account for multidimensionality rather than assuming binary categories or population homogeneity.
- Without better data systems, we miss clinically important knowledge and risk drawing false or incomplete conclusions.

Table 1.

Example of an anatomical inventory. Each data field should link to a corresponding code from the International Statistical Classification of Diseases and Related Health Problems. The format of this anatomical inventory is designed for expansion to accommodate additional variables and emerging health priorities over time.

Breasts □ Present □ Absent

- □ Chest reconstruction
- □ Bilateral mastectomy
- □ Unilateral mastectomy, R
- □ Unilateral mastectomy, L
- □ Breast augmentation/implants

Uterus □ Present □ Absent

- □ Hysterectomy—cervix removed
- □ Hysterectomy—cervix remains

Ovaries □ Present □ Absent

- □ Bilateral salpingo-oophorectomy
- □ Unilateral salpingo-oophorectomy, R
- □ Unilateral salpingo-oophorectomy, L

Cervix □ Present □ Absent

Grasso C, Goldhammer H, Thompson J, Keuroghlian AS. Optimizing gender-affirming medical care through anatomical inventories, clinical decision support, and population health management in electronic health record systems. J Am Med Inform Assoc. 2021;28(11):2531-2535. doi:10.1093/jamia/ocab080

Case study

Designing a better study

A research team is using electronic health record (EHR) data from a large health system in the United States to study prostate cancer incidence and mortality. The study population is defined as “male patients ages 45+” based on the electronic health record sex field. Available variables include age, race/ethnicity, socioeconomic status, comorbidities, and treatment, but there is no measure of anatomical inventory (e.g., presence of a prostate) or gender identity.

In their preliminary analysis, the team finds that Black patients have higher prostate cancer mortality than White patients, even after adjustment for key clinical and demographic factors. The team also notes a small number of prostate cancer diagnoses among patients recorded as female, but excludes these because they do not meet the “male patient” eligibility criteria. The team concludes that the racial mortality disparity may reflect biological differences in disease aggressiveness and that the excluded cases are likely coding errors.

Discussion questions

- How did the researchers define the population at risk?
- What kinds of misclassification could be occurring?
- What are the potential health consequences of such misclassification?
- What variables would improve the study design?
- What forms of differential exposure might explain the racial mortality disparity?

References

Bauer GR. Sex and Gender Multidimensionality in Epidemiologic Research. *Am J Epidemiol*. 2023;192(1):122-132. doi:10.1093/aje/kwac173

Berner AM, O'Callaghan S, Jackson SS. Methodology Concerns Regarding Claims Data Studies in Transgender Health. *JAMA Oncol*. 2025;11(5):567-568. doi:10.1001/jamaoncol.2025.0023

Grasso C, Goldhammer H, Thompson J, Keuroghlian AS. Optimizing gender-affirming medical care through anatomical inventories, clinical decision support, and population health management in electronic health record systems. *J Am Med Inform Assoc*. 2021;28(11):2531-2535. doi:10.1093/jamia/ocab080

Hamnvik OPR, Dizon DS, Leone AG. Methodology Concerns Regarding Claims Data Studies in Transgender Health. *JAMA Oncol*. 2025;11(5):567. doi:10.1001/jamaoncol.2025.0020

Krieger N. Genders, sexes, and health: what are the connections--and why does it matter? *Int J Epidemiol*. 2003;32(4):652-657. doi:10.1093/ije/dyg156

Manfredi C, Franco A, Ditonno F, et al. Prevalence and Factors Associated With Prostate Cancer Among Transgender Women. *JAMA Oncol*. 2024;10(12):1697-1700. doi:10.1001/jamaoncol.2024.4335

Who Counts in Prostate Cancer Epidemiology?: Rethinking Sex, Gender, & Risk

By Julia Naganuma-Carreras, Ashlynn Wimer, and Kris Berg

“What are these slides?”

These slides are meant to serve as templates for a course, with the goal of being hot swappable with already existent slideshows. Each slide has extensive speaker notes explaining the visual content of the slides. Feel free to use or remix as needed – everything produced here is under a Creative Commons license.

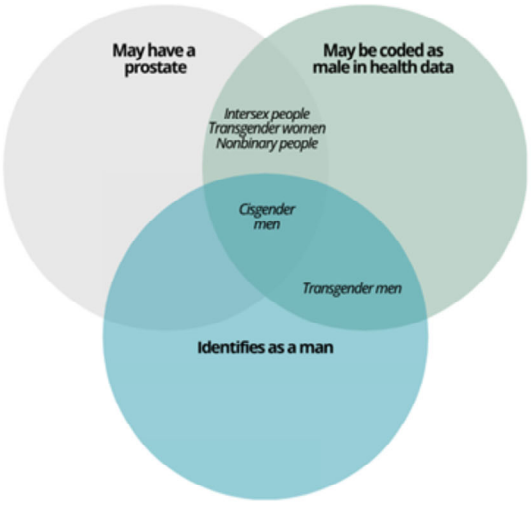
Who gets counted in cancer epidemiology?

Background

- Prostate cancer is often described as a cancer affecting "men"
- Epidemiologically, the population at risk is more precise: *peopl with prostates*
- "Men" is an imprecise proxy for prostate cancer risk
  - Some men, including many transmasculine people, *do not* have prostates
  - Some women, including many transfeminine people, *do* have prostates

Significance

- Incidence estimates depend on the correct denominator
- Screening depends on identifying who is actually at risk
- Studies relying only on binary sex/gender may exclude or misclassify relevant populations



Prostate cancer can also impact transgender women and non-binary people with prostates, and this group has unique risk factors when compared to men. For instance, transgender women may receive a variety of gender affirming surgeries, and often utilize endogenous hormones (gender affirming hormone therapy, or in this case feminizing hormone therapy). However, little is known about the epidemiology of prostate cancer among transgender women; this is largely attributable to challenges identifying transgender women (and transgender or non-binary people more broadly) within modern datasets, that often rely on a single binary gender or sex indicator.

To circumvent this, authors attempting to understand the epidemiology of cancer among this population often attempt to identify transgender people using billing codes data. Manfredi et al. (2024) is one such study, specifically seeking to understand prostate cancer outcomes among transgender people.

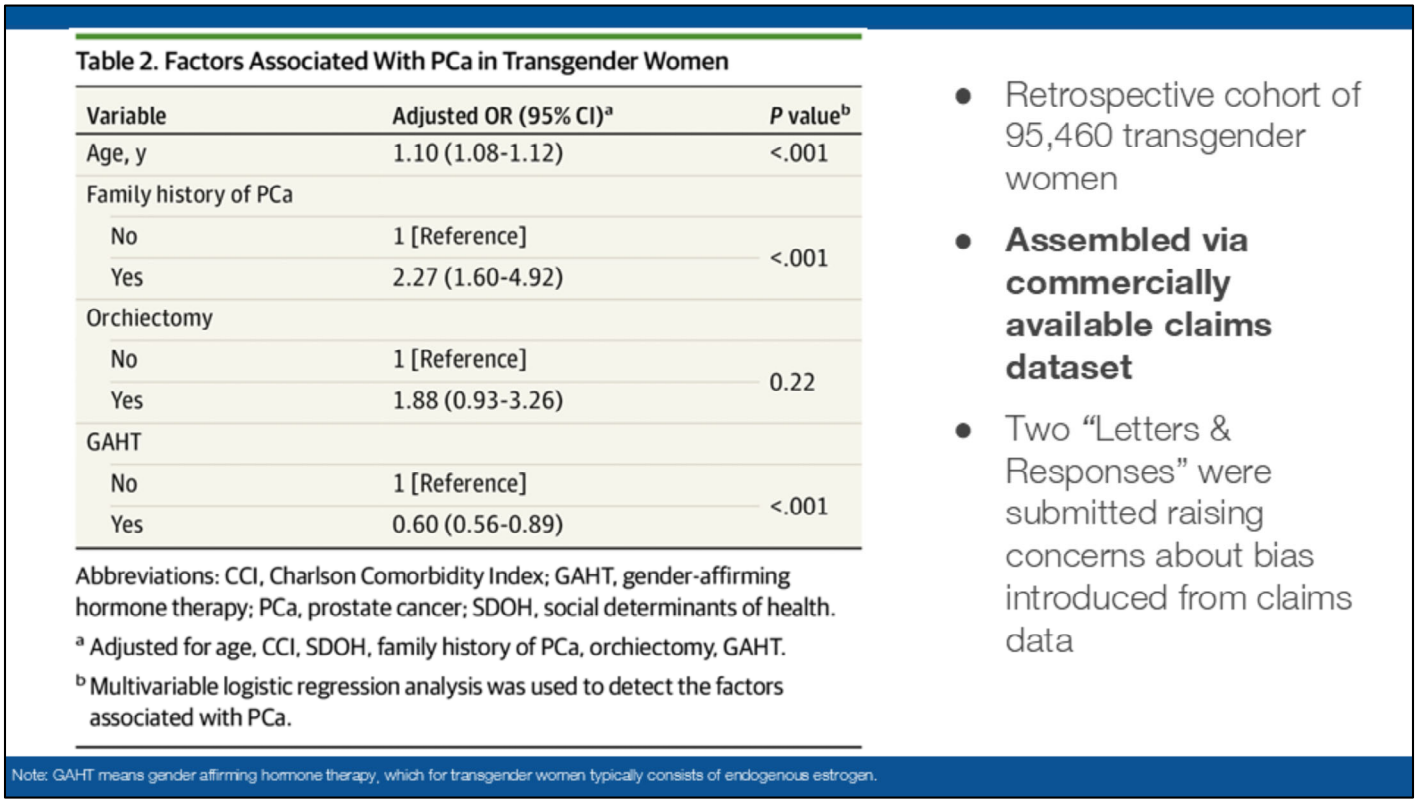


Table 2. Factors Associated With PCa in Transgender Women

Variable	Adjusted OR (95% CI)[a]	*P* value[b]
Age, y	1.10 (1.08-1.12)	<.001
Family history of PCa		
No	1 [Reference]	<.001
Yes	2.27 (1.60-4.92)	
Orchiectomy		
No	1 [Reference]	0.22
Yes	1.88 (0.93-3.26)	
GAHT		
No	1 [Reference]	<.001
Yes	0.60 (0.56-0.89)	

Abbreviations: CCI, Charlson Comorbidity Index; GAHT, gender-affirming hormone therapy; PCa, prostate cancer; SDOH, social determinants of health.

[a] Adjusted for age, CCI, SDOH, family history of PCa, orchiectomy, GAHT.

[b] Multivariable logistic regression analysis was used to detect the factors associated with PCa.

- Retrospective cohort of 95,460 transgender women
- **Assembled via commercially available claims dataset**
- Two "Letters & Responses" were submitted raising concerns about bias introduced from claims data

Note: GAHT means gender affirming hormone therapy, which for transgender women typically consists of endogenous estrogen.

Manfredi et al. (2024) found that gender affirming hormone therapy was protective of prostate cancer, that family history increased odds of the condition, and that orchiectomy was possibly a risk factor for prostate cancer.

However, methods were contentious: their cohort was assembled retroactively using a decade worth of deidentified claims data, an approach that prompted two "Letters & Responses" submitted about their article raising concerns about possible biases.

Briefly, both letters identified the very low percentage of participants receiving GAHT in the cohort Manfredi et al. (2025) assembled – 30.1% versus an expected ~70% – indicating a potential undercounting of transgender people and/or miscoding of cisgender people as transgender. Berner et al. (2025) even pointedly notes that, without a description of "how codes were used in combination to define the cohort", it was impossible to know if the authors had accidentally enrolled cisgender women in the cohort. Both letters also highlight that orchiectomy and GAHT, as-coded, are also at times *treatments* for prostate cancer, further increasing the classification inclusion.

The two letters:

Hamnvik OPR, Dizon DS, Leone AG. Methodology Concerns Regarding Claims Data Studies in Transgender Health. *JAMA Oncol*. 2025;11(5):567. doi:10.1001/jamaoncol.2025.0020

Berner AM, O'Callaghan S, Jackson SS. Methodology Concerns Regarding Claims Data

Studies in Transgender Health. *JAMA Oncol*. 2025;11(5):567-568.
doi:10.1001/jamaoncol.2025.0023

eTable 1. Codes used to identify the required data

	ICD-9	ICD-10	CPT	USC
Transsexualism	302.50, 302.51, 302.52, 302.53, 302.6, 302.85	F64.0, F64.1, F64.2, F64.8, F64.9, Z87.890	55970	/
Orchiectomy	62.41	0VTC0ZZ, 0VTC4ZZ	54520, 54690	/
Penectomy	64.3	0VTS0ZZ, 0VTSXZZ	54125	/
Vaginoplasty	64.5	0W4M070, 0W4M0J0, 0W4M0K0	57335, 57291, 57292, 56805, 58999, 56800	/
GAHT	V07.59, V07.51, V87.43	Z79.810, Z79.818, Z92.23	/	/
PCa	185, V10.46	C61, Z85.46	/	/
Family history of PCa	Z80.42	V1642	/	/
PSA	/	/	84153	/
BCR	/	R97.21	/	/
RM	198.5	C79.51	/	/

(Slide useful as an image when explaining what claims data is and what it's usage entails)

ICD-9, ICD-10, CPT, or USC "codes" to explain what specific conditions are diagnosed or treatments administered. Manfredi et al. (2024) required that at least one "transsexualism" code was present in patients previously marked as "male" when specifying their cohort.

Notably: while the transsexualism codes are accurately labeled (they all revolve around gender dysphoria (ICD-10, CPT) or "transsexualism" (ICD-9), the GAHT codes are somewhat inaccurately labeled. Specifically, they are:

- Z92.2: "Personal history of estrogen therapy"
- Z79.810: "Long term (current) use of selective estrogen receptor modulators (SERMs).
- Z79.818: "Long term (current) use of other agents affecting estrogen receptors and estrogen levels"
- V07.59: "Use of other agents affecting estrogen receptors and estrogen levels"
- V07.51: "Use of selective estrogen receptor modulators (SERMS)"
- V87.43: "Personal history of estrogen therapy"

What biases may be introduced through the usage of retrospective claims data?

Discussion is expected to revolve around various forms of **selection bias** and **misclassification,** as the cohort only consists of members whose treatments were accurately coded. Students may discuss risk of exposure misclassification (e.g. in dosage, or in terms of specific GAHT treatment), and of misclassification of “transsexual” status (e.g. if a transgender man does not out himself until *later* in his care). Later could introduce immortal time bias, as transgender men are not at risk of prostate cancer.

Potential biases that may be mentioned:

- Inclusion of **transgender men** who did not out themselves until *after* being registered as male in the databases. Would introduce people into the cohort who are not at risk of the outcome, and who would decline to receive either the GAHT specified or orchiectomy, making any association less protective than it should be.
- Potential misclassification of cisgender people (both sexes) as transgender. This could constitute misclassifying cisgender men or cisgender women as such, and in each case the specifics of the misclassification would impact the directionality.
- Reverse causation (given that orchiectomy and estrogen therapy are treatments for prostate cancer).
- Selection bias: the only people selected are those whose doctors accurately identified them as transgender, which is necessarily a crowd that is either earlier in transition (and hence *must* out themselves) or is later in transition and *willing* to reveal their transgender status.

Key takeaways

- Prostate cancer among transgender women highlights broader methodologic issues in cancer epidemiology.
- Study design and cohort selection can bias findings when “men” or “males” is used as a proxy for “people with prostates.”
- Better data collection should match what is being measured, including anatomy, gender identity, sex assigned at birth, hormones, and surgical history.
- Research should account for multidimensionality rather than assuming binary categories or population homogeneity.
- Without better data systems, we miss clinically important knowledge and risk drawing false or incomplete conclusions.

Table 1.

Example of an anatomical inventory. Each data field should link to a corresponding code from the International Statistical Classification of Diseases and Related Health Problems. The format of this anatomical inventory is designed for expansion to accommodate additional variables and emerging health priorities over time.

Breasts ☐ Present ☐ Absent
- ☐ Chest reconstruction
- ☐ Bilateral mastectomy
- ☐ Unilateral mastectomy, R
- ☐ Unilateral mastectomy, L
- ☐ Breast augmentation/implants

Uterus ☐ Present ☐ Absent
- ☐ Hysterectomy—cervix removed
- ☐ Hysterectomy—cervix remains

Ovaries ☐ Present ☐ Absent
- ☐ Bilateral salpingo-oophorectomy
- ☐ Unilateral salpingo-oophorectomy, R
- ☐ Unilateral salpingo-oophorectomy, L

Cervix ☐ Present ☐ Absent

Grasso C, Goldhammer H, Thompson J, Keuroghlian AS. Optimizing gender-affirming medical care through anatomical inventories, clinical decision support, and population health management in electronic health record systems. J Am Med Inform Assoc. 2021;28(11):2531-2535. doi:10.1093/jamia/ocab080

Case study

Designing a better study

A research team is using electronic health record (EHR) data from a large health system in the United States to study prostate cancer incidence and mortality. The study population is defined as "male patients ages 45+" based on the electronic health record sex field. Available variables include age, race/ethnicity, socioeconomic status, comorbidities, and treatment, but there is no measure of anatomical inventory (e.g., presence of a prostate) or gender identity.

In their preliminary analysis, the team finds that Black patients have higher prostate cancer mortality than White patients, even after adjustment for key clinical and demographic factors. The team also notes a small number of prostate cancer diagnoses among patients recorded as female, but excludes these because they do not meet the "male patient" eligibility criteria. The team concludes that the racial mortality disparity may reflect biological differences in disease aggressiveness and that the excluded cases are likely coding errors.

Discussion questions

- How did the researchers define the population at risk?
- What kinds of misclassification could be occurring?
- What are the potential health consequences of such misclassification?
- What variables would improve the study design?
- What forms of differential exposure might explain the racial mortality disparity?

References

Bauer GR. Sex and Gender Multidimensionality in Epidemiologic Research. *Am J Epidemiol*. 2023;192(1):122-132. doi:10.1093/aje/kwac173

Berner AM, O'Callaghan S, Jackson SS. Methodology Concerns Regarding Claims Data Studies in Transgender Health. *JAMA Oncol*. 2025;11(5):567-568. doi:10.1001/jamaoncol.2025.0023

Grasso C, Goldhammer H, Thompson J, Keuroghlian AS. Optimizing gender-affirming medical care through anatomical inventories, clinical decision support, and population health management in electronic health record systems. *J Am Med Inform Assoc*. 2021;28(11):2531-2535. doi:10.1093/jamia/ocab080

Hamnvik OPR, Dizon DS, Leone AG. Methodology Concerns Regarding Claims Data Studies in Transgender Health. *JAMA Oncol*. 2025;11(5):567. doi:10.1001/jamaoncol.2025.0020

Krieger N. Genders, sexes, and health: what are the connections--and why does it matter? *Int J Epidemiol*. 2003;32(4):652-657. doi:10.1093/ije/dyg156

Manfredi C, Franco A, Ditonno F, et al. Prevalence and Factors Associated With Prostate Cancer Among Transgender Women. *JAMA Oncol*. 2024;10(12):1697-1700. doi:10.1001/jamaoncol.2024.4335