**Supplemental Methods**

**Fewer than 0.5% of US Clinical Trials Enroll Pregnant Participants**

Alyssa Bilinski, PhD\*,1,2 · Natalia Emanuel, PhD\*3

\* Bilinski & Emanuel contributed equally.

1 Department of Health Services, Policy, and Practice, Brown University School of Public Health, Providence, Rhode Island

2 Department of Biostatistics, Brown University School of Public Health, Providence, Rhode Island

3 Federal Reserve Bank of New York, New York, New York

# A Data

We extracted data on April 1, 2024 from the Aggregate Analysis of Clinical Trials database, which compiles trials from ClinicalTrials.gov. There were 109,053 randomized controlled trials (RCTs) of drugs posted on ClinicalTrials.gov. We excluded RCTs with enrollment gender designated as only male (n=6,322), a minimum age above 45 (n=4,253), or a maximum age below 18 (n=5,220), leaving 93,632 RCTs. To restrict to RCTs likely required to register on ClinicalTrials.gov, we limited the sample to RCTs with at least one location in the US, US government funding, or industry funding, leaving 60,474 RCTs. Of these, we further excluded RCTs with "Unknown status" (n=2,778) and those that were "Withdrawn" prior to enrollment (n=2,367). We then reviewed (see "Other Classifications" prompt below) 655 RCTs with formal age or sex fields denoted as missing and excluded an additional 146 RCTs from which women aged 18-45 were excluded (e.g., enrolling excluding infants, children, post-menopausal women) or most likely excluded (e.g., focused on Alzheimer’s disease). This left us with a sample of 55,183 RCTs enrolling women aged 18-45. We later restricted our sample to RCTs posted between 2008 (when registration became mandatory) and 2023, 44,160 RCTs. We extracted fields for title, status, age, gender, type, masking, and funding sources as well as free text fields for summary, inclusion/exclusion criteria, and conditions.

# B Pregnant Inclusion

## II.A Training

To classify pregnant inclusion, we first manually labeled a training set (n=947); i.e. a human read each title, summary, and inclusion criteria and determined whether pregnant people were included, excluded, or not mentioned in trial criteria. We then evaluated the performance of GPT-4-Turbo-Preview (with prompt "Pregnancy Inclusion, Training") on this training set. We identified 22 errors related to pregnant inclusion (2.3%), with 3 errors among 170 RCTs labeled as included (1.8%), 14 among 470 RCTs labeled excluded (2.9%), and 5 among 308 labeled as unspecified (1.6%). We found no errors related to abortion/fetal demise classification, with 4 studies identified.

With these results and the release of GPT-4o, we modified our prompt slightly ("Pregnancy Inclusion, Full Sample") and evaluated performance on this same training set, with 14 errors (1.5%). We identified 13 errors (2.7%) among 482 RCTs labeled excluded and 1 error among 294 RCTs labeled as unspecified (0.3%). In the latter case, the main error identified was unspecified RCTs classified as excluding pregnant participants. To address this, we ran a secondary prompt ("Pregnancy Inclusion Check"), that prompted GPT to review output and verify whether its classifications were correct. With this prompt, GPT identified 7 errors (50%), without introducing any new errors, reducing errors on excluded-labeled RCTs to 7 of 482 RCTs (1.4%). We again found no errors related to abortion/fetal demise classification, with 4 studies identified.

## II.B Full Analysis and Validation

We ran our "Pregnancy Inclusion, Full Sample" and "Pregnancy Inclusion Check" prompts on the full sample of RCTs from 2008-2023. As an additional step, a human RA developed code-based text mining to classify RCTs. We manually reviewed all disagreements related to included RCTs and a subset from other classifications, noting that >80% of disagreements resolved in favor of GPT-based classification. From this we reclassified 1 trial as included, 3 as excluded, and 15 RCTs as unspecified (from excluded).

In this pool, after checking and removing RCTs related to abortion/treatment of fetal demise and postpartum conditions (see below), we reviewed a set of additional RCTs for validation. Given their importance to the study, we manually reviewed all

365 RCTs marked as included, identifying 3 that did not enroll pregnant participants (1.5% error rate on non-training RCTs). We also manually reviewed a random sample of 100 RCTs marked as excluded, noting 2 errors (both correctly reclassified as unspecified), and a random sample of RCTs marked as unspecified, noting 1 error (reclassified as excluded). Scaled to the distribution of identified RCTs, this suggested a 1.8% overall misclassification rate in the analysis. Importantly, however, we identified no errors related to missed "included" RCTs among "excluded" or "unspecified" RCTs in our review, and our manual review included all RCTs marked included to check for false positive classifications.

# C Other Classifications

We then ran additional prompts to identify other trial subsets, manually reviewing output to assess whether RCTs had been appropriately categorized.

* **Abortion/Treatment of fetal demise:** In our training set (above), we identified no errors in RCTs related to abortion or treatment of fetal demise. (This category did not include any RCTs focused on preventing miscarriage, only intervening after demise.) In the full sample, the prompt "Pregnancy Inclusion, Full Sample" flagged 84 RCTs, and after manual review of these, all were deemed relevant.
* **Postpartum interventions:** We used both "Pregnancy Inclusion, Full Sample" and "Postpartum Breakdown" prompts to identify RCTs delivering interventions in the postpartum period. After review of 76 RCTs, we identified 46 that enrolled pregnant individuals but delivered interventions only postpartum. In this category, we included all RCTs treating postpartum Cesarean pain (n=67) due to its postpartum focus and frequent ambiguity about treatment timing relative to delivery. We further reviewed all "included" RCTs manually for postpartum-only interventions (see above) and removed 1 accordingly.
* **Erectile dysfunction:** We used the prompt "Erectile Dysfunction" to evaluate randomized drug RCTs related to treatment of erectile dysfunction on the subset of RCTs enrolling only male participants from 2008-2023. This identified 116 RCTs, all of which were manually reviewed and deemed relevant.
* **Infertility:** We used the prompts "Fertility" and "Fertility, Male" (on the subset of randomized drug RCTs enrolling men) to evaluate randomized drug RCTs related to infertility treatment from 2008-2023. These identified 232 RCTs, which were all manually reviewed, and of which 225 were deemed relevant.
* **Conditions studied:** We categorized conditions studied based on the "Categories" prompt below and manually reviewed all identified studies for accurate categorization.
* **Missing age/sex:** For RCTs with age or sex fields marked "Missing", we used the "Missing age/sex", manually reviewing all RCTs indicated accordingly to exclude all women 18-45.

# D Prompts

We developed prompts by exploring performance on 25-50 RCTs, with a particular focus on improving performance by reviewing GPT’s summary providing reasoning for its classifications.

* **Pregnancy Inclusion, Training, GPT-4-Turbo-Preview:** "You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) a variable called AnyPregGPT indicating status of pregnant individuals in the trial. This can take one of 3 values. a) Unspecified: By default, mark a study Unspecified if pregnant individuals were not mentioned in the inclusion or exclusion criteria and/or the trial does not specify inclusion or exclusion based on pregnancy status – e.g., if pregnancy/lactating/contraceptives/childbearing were not mentioned in inclusion or exclusion criteria. b) Included: If and only if pregnant people could explicitly meet inclusion criteria for the clinical trial, mark this field as Included. c) Excluded: If and only if pregnant/lactating people (or in the pregnant stage) were explicitly excluded from the clinical trial (including by stating participants must take contraceptives to participate, the study requires a negative pregnancy test, or the trial does not include participants aged 18-45 years), mark this field as Excluded. Only mark this field as Excluded based on explicit quotable text related to pregnancy in study description. Studies that only fail to specify inclusion should be marked as Unspecified. (iii) a variable called Preg\_Only\_GPT which a value of 1 if and only if the trial only enrolled pregnant people and 0 otherwise; (v) ’Nonviable\_GPT’ equal to 1 if the trial involves nonviable pregnancies/fetuses, abortion, fetal demise, miscarriage, ectopic pregnancy, dilation and evacuation, dilation and curettage, or termination; (vi) ’Postpartum\_GPT’ equal to 1 if and only if a trial enrolls postpartum individuals or is evaluating a treatment deployed to mothers/birthing parents only after birth or delivery in the postpartum period. If the drug or treatment is giving during pregnancy, labor, or delivery, mark this field as 0.; (vii) ’Child\_GPT’ equal to 1 if the trial enrolls pregnant people but is testing a drug or intervention that will be provided to children after birth; (viii) a variable called Summary containing a summary of why you made these classifications. (ix) Summary quote: If a AnyPregGPT is marked Included or Excluded, directly quote the text related to pregnancy that led to this classification. If you cannot, AnyPregGPT should be marked as Unspecified. (x) a variable called Nonviable\_quote: If a Nonviable is marked 1, directly quote the text that led to this classification and explain reasoning. (ix) a variable called Postpartum\_quote: If a Postpartum is marked 1, directly quote the text that led to this classification and explain reasoning."
* **Pregnancy Inclusion, Full Sample, GPT-4o:** "You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) a variable called AnyPregGPT indicating status of pregnant individuals in the trial. This can take one of 3 values. a) Unspecified: By default, mark a study Unspecified if pregnant individuals were not mentioned in the inclusion or exclusion criteria and/or the trial does not specify inclusion or exclusion based on pregnancy status – e.g., if pregnancy/lactating/contraceptives/childbearing were not mentioned in inclusion or exclusion criteria. b) Included: If and only if pregnant people could explicitly meet inclusion criteria for the clinical trial, mark this field as Included. c) Excluded: If and only if pregnant/lactating people (or in the pregnant stage) were explicitly excluded from the clinical trial (including by stating participants must take contraceptives to participate, the study requires a negative pregnancy test, or the trial excludes participants aged 18-45 years), mark this field as Excluded. Only mark this field as Excluded based on explicit quotable text related to pregnancy in study description. Studies that only fail to specify inclusion should be marked as Unspecified. (iii) a variable called Preg\_Only\_GPT which a value of 1 if and only if the trial only enrolled pregnant people and 0 otherwise; (v) Nonviable\_GPT equal to 1 if the trial studies an intervention or medication used for nonviable pregnancies/fetuses, abortion, fetal demise, miscarriage, ectopic pregnancy, dilation and evacuation, dilation and curettage, or termination and 0 otherwise. This field should be 0 if cases like nonviable fetuses or fetal demise are excluded from the study.; (vi) Postpartum\_GPT equal to (a) Enrolls postpartum if it enrolls postpartum individual within 1 month of giving birth; (b) Postpartum intervention if it enrolls pregnant people BUT evaluates a treatment deployed to mothers/birthing parents only AFTER birth or delivery in the postpartum period; (c) otherwise is Neither. If the drug or treatment is given during pregnancy, labor, or delivery, mark this field as Neither.; (vii) ’Child\_GPT’ equal to 1 if the trial enrolls pregnant people but is testing a drug or intervention that will be provided to children after birth; (viii) a variable called Summary containing a summary of why you made these classifications. Remember that at trial should not be listed as Excluded on the basis of age unless it enrolls only neonates, infants, children, toddlers, perimenopausal women, or postmenopausal women. It can be listed as Excluded if it only enrolls these groups. If a study enrolls premenopausal women or includes any women aged 18-45, it should not be marked Excluded only on the basis of age. You can mark studies as Excluded if they study IVF, fertility treatments, hysterectomy, or contraceptives. Note any concerns about the AnyPregGPT classification. (ix) Summary quote: If a AnyPregGPT is marked Included or Excluded, directly quote the text related to pregnancy that led to this classification from the title, summary, or inclusion criteria. If you cannot, AnyPregGPT should be marked as Unspecified, and this field should say Unspecified. (x) a variable called Nonviable\_quote: If a Nonviable is marked 1, directly quote the text that led to this classification and explain reasoning. Otherwise, write 0. (ix) a variable called AnyPregGPT\_updated that updates AnyPregGPT based on the Summary quote field. If Summary quote is field lists Unspecified, this field should be listed as Unspecified. (ix) a variable called Postpartum\_quote: If a Postpartum is marked 1, directly quote the text that led to this classification and explain reasoning. Otherwise write 0."
* **Pregnancy Inclusion Check, GPT-4o:** You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) a variable called ’ChatGPT\_Check’. Read the justification and quote given, and state whether the evidence given indeed suggests pregnant/lactating people are excluded from the trial (Excluded), or whether the evidence given suggests that the trial did not specify whether pregnant/lactating people were included (Unspecified). (iii) Explain why you agree or disagree.
* **Erectile Dysfunction, GPT-4o:** You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) A variable Erectile equal to 1 if the trial studies an intervention related to erectile dysfunction/male impotence and 0 otherwise. Only mark this as 1 if the trial is focused on erectile dysfunction; if the trial is just studying a drug that can be used for erectile dysfunction for some other purpose, mark this as 0. (iii) A variable Erectile\_quote that provides the quoted text justifying this judgement if erectile = 1. (iv) A variable Erectile\_summary that explains the reason for this classification if Erectile is equal to 1.
* **Fertility, GPT-4o:** You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) A variable Fertility equal to 1 if the trial studies an intervention related to infertility or promoting fertility/conception, including in vitro fertilization (IVF), intrauterine insemination (IUI) or other assisted reproductive techniques and 0 otherwise. The study should enroll participants trying to conceive; if it enrolls pregnant individuals, this should be marked as 0. (iii) A variable Fertility\_quote that provides the quoted text justifying this judgement if fertility = 1. (iv) A variable Fertility\_summary that explains the reason for this classification if Fertility is equal to 1.
* **Fertility, Male, GPT-4o:** You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) A variable Fertility equal to 1 if the trial studies an intervention related to fertility/infertility/male factor infertility/conception (e.g., in vitro fertilization, intrauterine insemination, sperm count/motility, assisted reproductive techniques) and 0 otherwise. Only mark this field as 1 if the study is directly related to fertility or conception; if a study only requires that men use contraception to participate, mark this field as 0. (iii) A variable Fertility\_quote that provides the quoted text justifying this judgement if Fertility = 1. (iv) A variable Fertility\_summary that explains the reason for this classification if Fertility is equal to 1.
* **Postpartum Breakdown, GPT-4o:** You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) Enrolls\_postpartum, equal to 1 if the trial enrolls postpartum individuals less than two months after giving birth and 0 otherwise. If postpartum women are part of the exclusion criteria, this should be marked 0. (iii) a variable called Postpartum intervention equal to 1 if the trial studies a drug or intervention that is given AFTER delivery/birth during the postpartum period and 0 otherwise. If any part of the drug or intervention is delivered during pregnancy, labor, or delivery, mark this field as 0. (iv) A variable called C\_section\_pain that is equal to 1 if the trial studies pain management/analgesia related to C-sections/Caesarean sections or other post-Cesarean care; (v) A variable called Labor\_delivery\_pain if the trial studies pain management/analgesia related to labor and delivery (including C-sections) (vi) a variable called Postpartum\_quote: If any of the previous variables are marked 1, concisely and directly quote the text that led to this classification. (vii) Explain your reasoning for your classifications.
* **Categories, GPT-4o:** You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) a variable called AnyPregGPT indicating whether pregnant individuals can enroll in the trial (1) or 0 otherwise (iii) a variable called Preg\_Only\_GPT which a value of 1 if and only if the trial only enrolled pregnant people and 0 otherwise. (iii) Choose the bestfitting ’Category’. 1. Prevention of pre-term labor/premature birth/preterm delivery/PROM/pPROM/Cervical insufficiency/Cervical cerclage; 2. Hypertension/preeclampsia/eclampsia/HELLP syndrome; 3.HIV (Prevention of Mother-to-Child Transmission / PMTCT, NO malaria); 4. HIV (non-PMTCT, NO malaria, NO tuberculosis); 5. Malaria (with or without HIV); 6. Anemia; 7. Nutrition supplementation/Vitamin D/folic acid (DO NOT INCLUDE IRON OR ANEMIA); 8. Fetal growth/IUGR/Fetal Growth Retardation; 9. Diabetes (Type I or Type II ONLY); 10. Gestational diabetes; 11. Vaccination; 12. Asthma; 14. Hypothyroidism/Thyroid disease; 15. C-section/caesarean; 16. Labor induction/Labor management/Other labor and delivery (NOT C-SECTION, NOT CAESAREAN); 17. Opioid use disorder/Substance use disorder/OUD/SUD; 18. PCOS; 18. Fibroids; 19. Asthma; 20. Smoking cessation; 21. Other infectious disease prevention/treatment (non-vaccine); 22. Pregnancy headache/leg cramps; 23. GBS; 24. UTI; 25. Heartburn/GERD; 26.

COVID-19; 27. Nausea and vomiting during pregnancy/Morning sickness/Hyperemesis; 28. Postpartum sepsis/hemorrhage/complications; 29. Other - infant disease or injury; 30. Other - maternal/parental/pregnant person disease or injury; 31. Tuberculosis (with or without HIV); 32. Postpartum depression/postpartum psychosis; 33. Other mental health; 34. Abortion/Fetal Demise/Dilation and Evacuation/Dilation and Curettage Return the full category in a single variable – e.g., 33. Tuberculosis (with or without HIV). (iv) If a second category matches very well, list the full category in a column labeled ’Secondary category.’ Otherwise, return ’None’ for ’Secondary category.’ Add a variable (v) Pain that indicates with 1 if a trial evaluates a pain management drug or intervention and 0 otherwise. Create (vi) a variable called Summary containing a summary of why you made these classifications. (vii) Evaluate your certainty on each classification on a scale from 1-100.

* **Missing age/sex, GPT-4o:** You are a helpful assistant designed to output JSON. For each entry, create (i) a variable labeled NCT\_ID with the NCT ID, (ii) A variable labeled ’Any\_women’ that indicates whether the trial enrolls any women (=1) of any age or is 0 otherwise. If a trial excludes ONLY pregnant or breastfeeding women but allows women who are not pregnant or not breastfeeding to enroll, this field should still be marked as 1. This field should only reflect gender restrictions., (iii) A variable labeled ’Enrolls 18-45’ that indicates whether the trial includes ANY people aged 1845 (=1) or is 0 otherwise. For example, this criteria might be 0 if a trial only enrolls infants or only enrolls people who are postmenopausal. This field should only reflect age restrictions., (iv) A variable labeled ’Reason’ that indicates the reason for this classification.