

1 **Developing and piloting a preterm birth registry prioritizing social determinants of health**
2 **and patient-reported data**

3
4 **Authors:** Layla Joudeh^{1*}, BA; Jodi Stookey², PhD; Johanna Burch¹, MD; Rachel L. Kaplan¹
5 PhD, MPH; Sylvia Guendelman³, PhD, MSW; Aline Armstrong², MSN, RN, PHN; Rebecca
6 Jackson¹, MD

- 7
8 1. Department of Obstetrics, Gynecology, & Reproductive Sciences, School of Medicine,
9 University of California San Francisco, San Francisco, California USA.
10 2. Maternal, Child & Adolescent Health, San Francisco Department of Public Health, San
11 Francisco, California, USA.
12 3. Center of Excellence in Maternal, Child and Adolescent Health (MCAH) and Wallace
13 MCAH Research Center, School of Public Health, University of California Berkeley,
14 Berkeley, California, USA.

15
16 *Corresponding Author: Layla Joudeh—layla.joudeh@ucsf.edu; 550 16th St, San Francisco, CA
17 94158

18 *Disclosures:* The authors report no real or perceived vested interests that relate to this article that
19 could be construed as a conflict of interest.

20
21 *Funding:* This study was funded by the University of California, San Francisco (UCSF) Preterm
22 Birth Initiative. RJ and JS received the funding.

23
24 *Author contributions:* LJ was responsible for data curation, formal analysis, investigation,
25 methodology, project administration, visualization, validation, and writing the original draft. JS
26 was responsible for supervision, conceptualization, data curation, methodology, funding
27 acquisition, and reviewing and editing the original manuscript. JB was responsible for data
28 curation, methodology, investigation, and reviewing and editing the original manuscript. RK,
29 SG, and AA were responsible for methodology and review and editing the original draft. RJ was
30 responsible for supervision, conceptualization, data curation, formal analysis, funding
31 acquisition, methodology, visualization, validation and writing the original draft.

32
33 *Ethical considerations:* The University of California, San Francisco Institutional Review Board
34 approved the study (#17-21932). Written consent was obtained from all participants.

35 **Abstract:**

36 Biomedical factors and social determinants of health (SDH) affect preterm birth (PTB). Given
37 the complexity of PTB and the increasing rates in the United States, public datasets involving
38 multicomponent variables—from biomedical to structural—can motivate novel interventions to
39 address PTB in the US. The primary aim of this study was to develop a PTB registry based on
40 multi-modal data collection tools that prioritize biomedical and SDH data and identify PTB
41 phenotype (i.e. spontaneous labor, preterm premature rupture of membrane, or medically
42 indicated). The secondary aim of this study was to execute a pilot study to assess feasibility. This
43 study aimed to describe recruitment practices, assess data accessibility and concordance, and to
44 provide an example of how the registry can be used to generate hypothesis and analyze data. We
45 designed the registry through a conceptual model provided by the Dahlgren and Whitehead
46 model of SDH using validated questionnaires and literature on PTB. The registry included a
47 survey, interview, and medical and birth certificate abstraction. To pilot the registry, we
48 recruited 92 participants who delivered preterm, were admitted for risk of preterm delivery, or
49 delivered at term at an urban public hospital. Enrollment was most effective in-person and in the
50 postpartum period. Consent to different parts of the registry was similar regardless of when
51 participants were recruited. There was also a range of data concordance depending on the data
52 source and chosen variable. The registry’s PTB phenotype algorithm identified the correct PTB
53 phenotype 100% of the time. The example analysis demonstrated six unique SDH domains.
54 Participants who delivered preterm reported an average of 11 total stressors and 19.7 protective
55 items and 66% had a significant medical or obstetric comorbidity. Results of this study
56 demonstrate that a PTB registry is feasible and could help advance research to prevent PTB.

57 **INTRODUCTION:**

58 Despite medical advancements in the United States, preterm birth (PTB) rates have been
59 steadily increasing since 2014 (1), and non-Hispanic Black individuals have a 50% higher rate of
60 PTB than other racial and ethnic groups (2). PTB is multifactorial and involves social, structural,
61 environmental, genetic, and biomedical contributions (3). The World Health Organization
62 defines social determinants of health (SDH) as “non-medical factors that influence health
63 outcomes: the conditions in which people are born, grow, work, live, and age, and the wider set
64 of forces and systems shaping the conditions of daily life. These forces and systems include
65 economic policies and systems, development agendas, social norms, social policies and political
66 systems (4).” Decreasing the racial disparity in PTB requires focused efforts on understanding
67 SDH (5,6), including housing, safety, racism, childcare, and work; community and individual
68 support systems; and access to and trust in the medical system. Recent literature has pointed to
69 the need to study how medical, genetic, and SDH factors interact to contribute to PTB (6–8). For
70 example, Hong et al. (8) illustrated that in order to decrease the racial disparities in PTB, SDH
71 should be studied in conjunction with genomics data.

72 One way to effectively study multiple PTB factors simultaneously is through a
73 comprehensive, open-access data registry. Data registries are a unique platform to collect
74 comprehensive data; their large samples offer access to a wide range of clinicians and scientists
75 and can expedite the research process (9–11). However, to our knowledge, there is no central
76 database that focuses on bringing medical, obstetric, and SDH of pregnant people into the
77 research spotlight. Birth registries typically record medical and demographic information but
78 vary on the inclusion and detail of SDH data. For example, the March of Dimes Database for
79 PTB Research focuses on genomic, transcriptomic, immunological, and microbiome data. While

80 databases like these are helpful in designing interventions that could improve birth outcomes,
81 they lack SDH data (12). Electronic health records (EHR) have been shown to be inaccurate for
82 collection of SDH data in comparison with patient report (13–15). A recent study found that
83 incorporating self-reported SDH data into EHRs improved the accuracy of hospital admission
84 risk estimations (16).

85 PTB phenotype is also important to consider when designing a PTB registry. There are a
86 variety of potential mechanisms and causes of PTB. Accurate determination of PTB phenotype
87 (e.g. preterm prelabor rupture of membranes (PPROM), spontaneous preterm labor (SPTL), or
88 medically indicated) is critical (17,18), and PTB phenotype has been found to be inaccurate in
89 administrative databases (19–22).

90 Our study aimed to develop and pilot a PTB registry that incorporates biomedical and
91 SDH data. Within the registry, we also aimed to develop an algorithm that identifies PTB
92 phenotype using medical record data. We then used the registry in a pilot study to assess its
93 feasibility. The pilot study aimed to describe recruitment practices, to assess data accessibility
94 and concordance, and to provide an example of how the registry can be used to generate research
95 questions and analyze data. This registry and pilot study can serve as a framework for designing
96 and implementing a patient-centered PTB registry. A feasible, comprehensive registry can lead
97 to robust research at a local and national level that will help reduce PTB rates and disparities.

98 **METHODS:**

99 An interdisciplinary team including epidemiologists, social workers, public health nurses,
100 physicians, and health professional students designed a comprehensive PTB registry. The
101 registry included a survey, interview, and medical record and birth certificate abstraction.

102 *Designing the registry*

103 We created the four data collection tools—survey, interview, medical record abstraction,
104 and birth certificate abstraction—based on existing research on risks factors for PTB, using
105 validated instruments when possible, and modified them based on input from a preterm birth
106 community advisory board. We designed our instruments using the Dahlgren and Whitehead
107 model of the main social determinants of health, a social-ecological framework that considers
108 health as a function of individual lifestyle factors, social and community networks, and socio-
109 economic, cultural, and environmental conditions (23). Our goals in designing the instruments
110 were to: (1) develop targeted questions that focus on community-level and individual level
111 biomedical and SDH variables, (2) use multiple sources to collect data in order to compare data
112 consistency across the data sources., and (3) include self-reported data. Our instruments focused
113 on biomedical variables and community and individual level SDH. The data collection tools
114 were iteratively revised based on participant and researcher feedback to improve implementation.
115 In the following paragraphs, we will describe each data collection instrument.

116 The final survey had 73 questions and spanned community-level and individual-level
117 factors (S1 File). Given the survey was designed as a screening tool, to limit the length of the
118 survey, we chose overarching questions from multi-question validated scales that measure
119 common determinants of health, such as housing instability. The community-level factors were
120 housing, work, financial security, food security, discrimination, social services, care quality, and
121 care access. The individual-level factors were health, health-related behaviors, life stressors, trust
122 in health care system and provider, and resilience factors. We identified survey questions and
123 variables using validated questionnaires that included the 2005 LAMBS, 2017 SOLARS, 2012
124 PRAMS, 2010 LA HOPE, and 2014 MIHA questionnaires as well as the California Health
125 Interview Survey and San Francisco FIMR (24–30). Survey items included both adverse SDH

126 (stressors) (e.g., housing instability) and potentially positive factors (e.g., utilization of
127 community resources).

128 The interview was composed of open-ended questions about pregnancy experiences to
129 systematically capture qualitative data that might complement quantitative results from the
130 survey and medical record and birth certificate abstractions. The interview transcripts were
131 coded to match the SDH and individual level factor domains in the other data instruments. A
132 previous study showed how qualitative and survey data can be analyzed in conjunction with one
133 another (35).

134 The main domains of the medical record abstraction were medical and obstetric co-
135 morbidities, risk factors for preterm delivery, labor and delivery care and diagnoses, prenatal
136 care and diagnoses, lab values, mental health, and substance use (S3 File). Because the PTB
137 phenotype was not directly reported in the medical record, the medical record abstraction also
138 included an algorithm that collected data in order to identify the PTB phenotype (Fig. 3 and S3
139 File). We also included variables to monitor quality of care such as receipt of aspirin,
140 progesterone, and betamethasone. The birth certificate domains included prenatal care and labor
141 and delivery data, newborn baby medical information, and demographic information about the
142 parent(s) (S2 File).

143 ***Piloting the registry:***

144 From October 2017 to March 2019, we recruited pregnant or immediately postpartum
145 individuals at a single urban public hospital in three groups: (1) postpartum after a PTB ; (2)
146 during admission for PTL, PPRM, or other preterm medical indications (e.g. preterm
147 hypertensive disorders), or (3) during a labor and delivery triage visit for PTL evaluation (Fig 1).
148 We also recruited a comparison group of participants who were approached after a term birth

149 with no preterm admission or triage visit for PTL (Fig. 1). We enrolled individuals who were at
150 risk for PTB but had yet to deliver preterm to determine feasibility of enrollment to the registry
151 at different points in the hospitalization. The comparison term group was recruited to
152 demonstrate how a case-control study analyzing PTB could utilize both the PTB registry and a
153 control term group.

154 Eligible participants were 18 years of age or older, spoke English or Spanish, and had a
155 viable pregnancy beyond 24 weeks gestation. We used convenience sampling and attempted to
156 screen and enroll all pregnant patients who were eligible for the study during our recruiting time
157 period.

158 ***Recruiting for the pilot:***

159 We used the EHR to identify eligible participants and contacted them up to three times
160 within one week of the first hospital presentation to share study information. The term group was
161 recruited to achieve similar proportions of race and ethnicity, parity, and language as the preterm
162 groups. Contact was made in person when possible and otherwise via phone calls. To increase
163 autonomy in research procedures, participants were asked to separately consent to each data
164 collection instrument (i.e. survey, interview, medical record abstraction, and birth certificate
165 abstraction). Participants chose when and where to do the survey and/or interview.

166 Participants completed the survey on paper or on a tablet and had the option to complete
167 the survey alone, with partial assistance, or with a researcher reading every question-and-answer
168 choice. Interviews were completed by phone or in person, recorded, and professionally translated
169 and transcribed. Participants received a \$20 gift card for completing the survey and a \$30 gift
170 card for completing the interview. Medical record and birth certificate abstractions were
171 completed postpartum to ensure all labor and birth data were available.

172 ***Data management:***

173 All study data were stored in Research Electronic Data Capture (REDCap), a secure web
174 platform for building and managing databases. We assessed feasibility with multiple measures.
175 We collected data on enrollment practices (i.e. success rate of enrollment based on if a person
176 was approached in-person versus over the phone; success rate of enrollment based on whether a
177 person was in triage, antepartum, or postpartum). We also recorded time required to complete the
178 survey and medical record abstraction, and asked participants their satisfaction with time to
179 complete the study. We also compared data concordance between birth certificate, medical
180 record abstraction, and survey. We report percent concordance (agreement) and Cohen’s kappa
181 (36,37). Kappa represents the degree of agreement, adjusted for the amount of agreement that
182 could be expected due to chance alone and is commonly interpreted as a high level of agreement
183 (0.81-1.0), substantial (0.61-0.8), moderate (0.41-0.6) and fair/slight (≤ 0.40) (36,37).

184 For purposes of analysis, we characterized risk factors for PTB as major if they were
185 consistently associated with PTB across several studies at an OR above 1.3. These major risk
186 factors include prior preterm birth, multiple gestation, short cervix, hypertensive disorders, renal
187 and cardiovascular disease, diabetes, cholestasis, “Black,” “other race”, BMI <19 , age >40 years,
188 tobacco use, and placenta previa or abruption (31–34). We also examined a fuller set of potential
189 risk factors that additionally included substance use disorder, depression, birth spacing <18
190 months, infection, intrauterine growth restriction, oligo- or polyhydramnios and third trimester
191 bleeding.

192 To illustrate one possible analysis that can be done using data from the registry, we
193 examined medical and obstetric comorbidities and SDH in three groups: (1) PTB; (2) term births
194 in those who were enrolled during a preterm labor (PTL) triage visit or admission for preterm

195 medical indications, such as PTL, PPRM, or hypertensive disorders; and (3) term births
196 without prior antepartum admission. We categorized SDH variables from the survey, medical
197 record, and birth certificate as potentially protective or as a perceived stressor and grouped them
198 into financial (including housing and food insecurity), employment, health care access and
199 experience, social, health, and personal domains (S1 Appendix). The number of items in each
200 domain ranged from two for employment protective factors to 11 for social stressors (S1
201 Appendix). Because there were differing numbers of variables in each domain, we normalized on
202 a scale of zero to 10 points for each domain. The average number of stressors and protective
203 factors per person was then calculated for each domain. We evaluated the bivariate association of
204 demographic, medical and SDH with birth group (PTB, term birth after PTL or antepartum
205 admission, term birth with no prior antepartum admission) using Fisher's exact tests for
206 categorical variables and t-tests for continuous variables. We considered p values less than 0.05
207 to be statistically significant. We performed all statistical analyses using Stata version 14.1 (38).

208 **RESULTS**

209 In this section, we first describe the sample characteristics for the registry's pilot testing.
210 Then we discuss the feasibility of the registry in terms of recruitment, data accessibility and
211 concordance, PTB phenotype. In the final section of the results, we illustrate an example data
212 analysis using data from the registry's pilot testing.

213 ***Sample characteristics for registry's pilot testing***

214 We enrolled a total of 92 participants from October 2017 to March 2019 (Fig 1) to pilot
215 test registry. Participants were enrolled prior to delivery (n=29) or after a preterm (n=43) or term
216 delivery (n=20). However, four of the participants recruited in triage and four of the participants
217 recruited antepartum did not consent to sharing their birth outcomes, so their data is excluded

218 from all subsequent analysis. For those enrolled prior to delivery, 11 were enrolled during an
219 antepartum admission for PPRM, PTL or other medical indications and six of the 11 (55%)
220 went on to deliver preterm; 10 were enrolled during a triage visit for PTL, and one delivered
221 preterm (10%) (Fig 1). Of 50 preterm births, 2 (4%) were ≤ 28 weeks gestation, 10 (20%) were
222 29-33 weeks and 38 (76%) were 34-27 weeks.

223 ***Best practices for recruitment***

224 Recruitment methods were analyzed to determine best practices for an opt-in registry.
225 Enrollment was most effective when participants were approached in person. 55% (n=78) of
226 individuals who were initially approached in person enrolled in the study, while only 21%
227 (n=13) of individuals who were initially approached over the phone enrolled in the study.
228 Comparing antepartum and triage to postpartum enrollment (excluding the comparison term
229 group who were only approached in postpartum), 46% were successfully enrolled in the
230 postpartum period compared with 32% during antepartum admissions and in triage (p=0.04). The
231 average time to complete the survey was 29 minutes (SD=13; range=10-80 minutes), and 94% of
232 survey participants felt it was just the right length. Of participants who completed the survey
233 (n=87), 28% (n=25) asked that the survey be fully read to them and 9% (n=8) had partial
234 assistance. The average time to complete medical record abstraction was 38 minutes (SD=14;
235 range=15-60 minutes).

236 ***Data accessibility and concordance***

237 To facilitate participant autonomy, consent was obtained separately for each part of the
238 registry (i.e. the survey, interview, prenatal records, labor and delivery records, birth certificate
239 data, and the HIPAA-specific topics of mental health, HIV, and substance use). Only 35%
240 consented to all portions of the study. 95% of participants consented to the survey. 82% of

241 participants consented to the interview. 57% of participants consented to obtaining data from the
242 birth certificate. 52% of participants consented to a full medical record abstraction. There were
243 no significant consent pattern differences among those recruited in triage or antepartum,
244 postpartum after a PTB, or postpartum after a term birth. An additional 22% of participants
245 consented to a nearly full medical record abstraction but declined either HIV status, substance
246 use history, and/or mental health information. 26% of participants declined either their complete
247 labor and delivery record and/or prenatal record.

248 Consistency of information was compared across birth certificate, medical record, and the
249 survey for select variables. Compared to medical record, the birth certificate was found to have
250 >98% agreement and kappa > 0.75 for mode of delivery, chronic hypertension, gestational age at
251 delivery, and induction of labor. Agreement was lower (76-92% concordance and kappa <0.70)
252 for other variables such as prior preterm delivery, PPRM, and use of betamethasone. The birth
253 certificate had both false positives and false negatives. False positives, in which a
254 condition/experience was reported in birth certificate but not confirmed in the medical record,
255 occurred for PPRM (15%) and first prenatal visit ≤ 8 completed weeks (15%). False negatives,
256 in which the condition was missed in the birth certificate but included in the medical record,
257 occurred for prior PTB (8.5%), first prenatal visit ≤ 8 completed weeks (15%), diabetes mellitus
258 (2%), PPRM (10%), IOL (3%), use of betamethasone (33%), and hypertension in pregnancy
259 (6%).

260 There were several differences noted when comparing medical record to survey data.
261 Depression had a very low kappa of 0.14 and a concordance of 67%, with 30% more participants
262 reporting depressive symptoms during their current pregnancy on the survey than carrying a
263 diagnosis of depression in the medical record. On the other hand, 25% of those with a diagnosis

264 of depression in the medical record did not report depressive symptoms during their current
265 pregnancy on the survey. Diagnosis of depression includes both current depression and history of
266 depression. Similarly, unstable housing was reported by 44% in the survey but recorded for only
267 29% in the medical record (78% concordance, kappa=0.56). Conversely, intimate partner
268 violence was more likely to be reported in the medical record (14%) compared with the survey
269 (7%) (90% concordance, kappa 0.45).

270 ***Preterm Birth Phenotype***

271 Based on comprehensive medical record abstraction, among PTBs, 30% occurred after
272 PPRM, 38% after spontaneous labor, and 32% were medically-indicated (62% of these for
273 hypertensive disorders). Using variables easily accessible in the medical record, we created an
274 algorithm using PPRM (Y/N), induction of labor (Y/N), and Cesarean plus medical conditions
275 known to be associated with indicated preterm delivery (Y/N) (Fig. 3 and S3 File). This
276 algorithm initially misclassified four (8%) of the PTBs labelling them as medically indicated
277 instead of spontaneous labor. The algorithm was revised to include a variable for PTL symptoms
278 (contractions, bleeding) at time of presentation, which was obtained from the history of present
279 illness field and the admission history and physical in the medical record. With this revision, the
280 algorithm correctly categorized all PTBs, as compared with comprehensive medical record
281 review. On the other hand, using only variables available in the birth certificate, 17% of births
282 were misclassified using the algorithm (kappa=0.74).

283 ***Example analysis***

284 A secondary aim of the study was to show how the registry can be used to generate
285 hypotheses and analyze data. In the subsequent section, we show one way to group and describe
286 study participants. Then we present how the registry's survey, medical record, and birth

287 certificate data can be organized and used for comparative analysis. In our sample pilot, the
 288 majority of participants identified as Latinx (57%) or Black/African American (24%), were
 289 publicly insured (96%), and had household income <\$39,000 (56%) (Table 1). 12% of
 290 participants had a prior PTB. 33% of participants had hypertensive disorder(s). 14% of
 291 participants had pre-existing or gestational diabetes during the pregnancy (Table 2).

292 **Table 1: Selected participant characteristics according to when recruited and delivered, n**
 293 **(%)**
 294

	Recruited preterm, Preterm birth n=50	Recruited preterm, Term birth n=14	Recruited term, Term birth n=20
Age, mean (SD)	31.0 (6.5)	32.4 (4.6)	31.4 (5.5)
Race/Ethnicity			
Latinx/Hispanic	30 (60)	5 (36)	13 (65)
Black/African American	11 (22)	6 (43)	3 (15)
White	1 (2)	2 (14)	1 (5)
Asian	1 (2)	0 (0)	1 (5)
Pacific Islander	1 (2)	0 (0)	0 (0)
Other/Unknown	6 (12)	1 (7.1)	2 (10)
Study completed in Spanish	21 (42)	3 (21)	9 (45)
Annual family income			
<\$16,000	16(48)	7 (58)	7 (54)
\$16-39,000	13(39)	3 (25)	1 (7.7)
>\$39,000	4 (12)	2 (17)	5 (39)
Employed during pregnancy	29 (62)	9 (69)	15 (83)
Housing instability*	13 (27)	6 (46)	2 (10)
Food insecurity^	17 (36)	8 (67)	4 (22)
Did not complete high school	10/34 (29)	3/8 (38)	4/9 (44)
Needs assistance to complete medical forms	9/48 (19)	1/13(7.7)	4/18 (22)
Nulliparous	15 (31)	2 (15)	8 (47)
First prenatal visit by 8 weeks of gestation	16/40 (40)	2/13 (15)	5/14 (36)
<1 day/week of exercise during pregnancy	32 (68)	3 (23)	5 (28)
Has someone who can help with daily tasks	41 (85)	10 (77)	13 (72)
Has someone who can provide emotional support	43 (90)	12 (92)	16 (89)

295

296 *Housing insecurity includes houseless or living in shelter, on couch or in SRO; or lives in place without kitchen or
 297 private bathroom.
 298 ^Food insecurity includes worrying about running out of food or use of food pantry.

299
 300 **Table 2: Selected risk factors for preterm birth according to when recruited and delivered,**
 301 **n (%)**
 302

	Recruited preterm, Preterm birth n=50	Recruited preterm, Term birth n=14	Recruited term, term birth n=20	p
Prior PTB	8 (19)	2 (14)	0 (0)	0.13
Obstetric/medical comorbidity[^]	27 (66)	10 (71)	5 (29)	0.02
HTN (chronic, gestational, or pre-eclampsia)	18 (44)	5 (36)	3 (21)	0.35
Depression	16 (32)	6 (43)	8 (40)	0.69
Substance use during pregnancy (excluding marijuana)	6 (21)	5 (42)	2(14)	0.28
Tobacco use during pregnancy	5 (12)	5 (36)	0 (0)	0.014
Black or race defined as “other”	17 (34)	7 (50)	5 (25)	0.34
Any major PTD risk factor*	33 (79)	10 (71)	6 (33)	0.003
Number of major PTD risk factors*				0.009
• 0	9 (21)	4 (29)	12 (67)	
• 1	12 (29)	3 (21)	4 (22)	
• 2-5	21 (50)	7 (50)	2 (10)	
Any PTD risk factor[#]	40 (91)	14 (100)	10 (56)	0.001

303
 304 [^] Hypertension; diabetes; renal, heart or liver disease; cholestasis; sickle cell disease; seizure disorder; hepatitis C; human immunodeficiency virus (HIV); pyelonephritis; admission for abscess or cellulitis; multiple gestation; placenta previa; short cervix; intrauterine growth restriction; oligohydramnios

305
 306 *Major PTD risk factors: risk factors with OR>1.3 in multiple studies, see methods section. Prior PTB, multiple gestation, short cervix, hypertensive disease, Black or other race, age > 40 yo, BMI<19, tobacco use, diabetes, heart or renal disease, placenta previa or abruption.

307
 308 # PTD risk factors include major PTD risk factors plus additional risk factors inconsistently shown in multiple studies: birth spacing <18 months, substance use, depression, asthma, infection (urinary, STI), intrauterine growth restriction (IUGR), oligo/polyhydramnios, third trimester bleeding.

309
 310
 311 In this section, we illustrate one way to use the registry data to analyze perceived
 312 protective factors and stressors that could affect PTB. The frequencies of individual protective
 313 factors (n=36) and stressor (n=40) SDH items are shown in S1 Appendix . The most commonly
 314 reported stressors among individuals who delivered preterm were: (1) obstetric or medical
 315 comorbidities (66%), (2) strenuous work conditions (45.8%), and (3) experiencing poverty as a

320 child (38.3%). Among individual items, the most commonly reported protective factors among
321 individuals who delivered preterm were: (1) during my pregnancy, I had someone I could turn to
322 if I needed someone to comfort me or listen to me (89.6%), (2) during my pregnancy, I had
323 someone to help me with daily tasks (85.4%), and (3) I am satisfied with my prenatal care
324 (81.3%). There were no statistically significant associations found between individual social
325 determinants and delivery group.

326 The mean number of SDH stressors and protective factors between the three study groups
327 are summarized in Fig. 2. There were no statistically significant differences across delivery
328 groups for each domain or for the total across all domains. Among stressors, the domain with the
329 highest mean score was work (2.84 out of 10). Among protective factors, the highest was health
330 care experience (5.3 out of 10). Participants who delivered preterm reported an average of 11
331 total SDH stressors and 19.7 total protective factor items of 60 possible (Fig 2). Although not
332 statistically significant, the group that delivered at term but was recruited preterm during
333 admission or triage reported more stressors in most categories than the other two groups.

334 *Individual-level risk factors*

335 Pregnant participants who delivered preterm and those who were recruited preterm but
336 delivered at term had significantly more major preterm delivery risk factors and medical co-
337 morbidities than those recruited at time of a term delivery ($p < .05$) (Table 2). Among those who
338 had a PTB, 18% had prior PTB, 44% had hypertensive disorder, and 23% had diabetes. There
339 was also significantly more tobacco use during pregnancy ($p = .009$), obstetrics or medical
340 comorbidity ($p = .02$), and at least one PTD risk factor ($p = .001$) in the pregnant people who
341 delivered preterm and those who recruited preterm but delivered at term compared to the
342 participants recruited at time of term delivery.

343 **DISCUSSION**

344 Our PTB registry pilot highlights methods to develop a feasible PTB registry that
345 captures SDH as well as biomedical data. We also demonstrate how data can be organized and
346 analyzed using a registry. Based on our study, we recommend the following best practices:

347 *Recruitment practices*

348 For recruiting participants, we recommend in-person, language-concordant recruitment to
349 adequately explain the purpose of the registry, obtain consent, and assist with completion of data
350 collection. One option for public health departments to consider would be to utilize birth
351 certificate clerks to complete registry procedures as they already visit each patient to collect birth
352 certificate data. We also recommend compensating participants for their time and information, as
353 this helps recruitment and engagement and acknowledges the contribution participants have
354 made to medical and public health advancements (39).

355 While in-person recruitment was effective in our sample, we still encountered varying
356 degrees of consent to specific study procedures. The majority of the participants completed the
357 survey, a data collection tool in which participants can control their responses. In contrast, fewer
358 participants consented to a birth certificate and full medical record abstraction. We were still able
359 to recruit participants using this opt-in process, but studies have shown that opt-out registries
360 yield higher participation (40). The challenge with opt-in and opt-out registries is that varying
361 degrees of consent yields missing data for primary and secondary analysis. However, given
362 mistrust in the medical system, especially in communities of color due to current and historical
363 injustices and discrimination (41–43), it is important to have transparent study procedures such
364 that participants know what data is being collected and how it is used. Thus, a priority for
365 registry design is to have a consensus about the key primary outcome measures that must be

366 collected for all registry participants (10). This will limit data gaps and ensure that when
367 individuals do participate in the registry, their information will contribute to robust multi-site
368 analysis while still facilitating autonomy over their information.

369 An additional aspect of registry implementation is consensus on what sample to include.
370 Table 2 shows that participants who presented to the hospital with PTB symptoms or conditions
371 that commonly lead to PTB and went on to deliver at term were more similar to those with PTB
372 than with term birth recruited at term. This group included both patients admitted with high-risk
373 conditions (PTL, hypertensive disorders) and patients seen in labor and delivery triage who had
374 PTL symptoms but were not admitted. It was more effective to enroll the admitted patients
375 compared with the triage patients, and there was a higher rate of PTB in the admitted versus the
376 triage patients (40% versus 7%). Including this group in a registry would therefore have the
377 benefit of illuminating the difference between these two groups with different birth outcomes
378 despite having similar symptoms and risk factors for PTB. While including this group would
379 involve more nuanced participant screening, it has the potential of providing more robust data.
380 We also found that participants were more likely to consent to participate if they were
381 approached in the postpartum period, so focusing participant recruitment during this time period
382 could be most effective in optimizing registry participation.

383 ***Data inclusion and analysis***

384 Our results illustrated that self-reported SDH data can be a more reliable way to include
385 SDH data in a registry, as only very limited SDH data is found in the medical record or birth
386 certificate. For the few SDH variables that were available in the medical record, like unstable
387 housing, the medical record under-reported or over-simplified this data. This aligns with
388 previous research that showed that medical records do not accurately capture SDH data (13–15).

389 While medical record abstraction was less accurate for SDH data, it was more accurate than the
390 birth certificate for the majority of medical and obstetric variables, similar to other studies
391 (22,44). Survey and medical record data were complementary in capturing sensitive medical
392 information. Including both medical record and self-reported data in a comprehensive registry
393 would enable researchers to investigate a variety of hypotheses. For example, using our registry
394 data, Reno et al. (35) demonstrated pregnant people receiving behavioral health services had
395 significantly later gestational ages at birth compared to participants in our sample who did not
396 receive behavioral health services.

397 A registry should also include sufficient data to accurately determine the PTB phenotype,
398 given that PTB is a heterogenous outcome (17,18). While there are alternative options for
399 classifying PTB, the most commonly used is PPROM, SPTL, and other medical indication
400 (17,18,45). Where possible, we recommend updating the EHR to include a drop-down menu that
401 enables streamlined PTB classification. In the absence of classification by the delivery provider,
402 we recommend determining PTB phenotype with EHR data or by using an algorithm (Fig. 3 and
403 S3 File), with inclusion of signs of spontaneous labor (contractions, bleeding) for improved
404 accuracy. We recommend against using birth certificate data to determine PTB phenotype and
405 risk factors associated with PTB. Compared with medical record abstraction, our birth certificate
406 data was misclassified phenotype in 17% and under- or over-reported on a number of other
407 variables. Studies comparing birth certificate data to medical record review similarly show that
408 birth certificate data do not accurately differentiate between PPROM, SPTL, and medically
409 indicated delivery, and under-report medical and obstetric complications (19,46,47).

410 Existing studies highlight the need for SDH data in understanding PTB, developing PTB
411 interventions, and improving birth outcomes. Our data analysis illustrates one of many possible

412 analyses using the registry data. A registry comparable to the one described here with large
413 sample size could provide the statistical power necessary to better understand the significant
414 relationships between SDH and preterm birth. Finally, example data analysis demonstrates how a
415 term group could be separately recruited from a PTB registry in order to use the registry to
416 conduct a case-control study. The data could also be adapted for quality improvement purposes.
417 For example, the data could be used to assess and ensure appropriate use of betamethasone and
418 aspirin.

419 ***Future studies***

420 While we have shown that there is a feasible framework for a comprehensive PTB
421 registry, implementing a registry requires logistical support, coordination, and acceptability
422 among health care providers and systems. A future study could assess health care providers'
423 desires to have a PTB registry and the acceptance of a PTB registry similar to the one presented.
424 Future directions could include using a large sample to conduct robust research that uniquely
425 examines interactions between clinical diagnoses, physical and mental health, SDH, and birth
426 outcomes. Finally, a financial assessment of registry development and maintenance would
427 contribute to the feasibility data provided by these results.

428 ***Limitations***

429 This pilot used a small sample to assess the feasibility. We also did not have the capacity
430 to include participants who spoke languages other than Spanish or English. These limitations raise
431 the possibility of selection bias, as the factors making potential participants unreachable,
432 ineligible, or unwilling to participate could be associated with risk for PTB. The information
433 gathered from these participants and their feedback on how to make the registry more accessible
434 and acceptable are especially important. Lastly, because this study took place at a single site, we

435 cannot generalize the results to other settings. However, despite these limitations, the results
436 described here provide a promising way forward for developing a PTB registry.

437 **CONCLUSION**

438 A PTB registry incorporating SDH with medical and obstetric factors is feasible.
439 Recruiting participants in person and in the postpartum period, having personnel to support
440 participants in completion of survey, leveraging complementary data sources including self-report
441 and medical records, ensuring transparency in data collection, and providing a participant consent
442 process that promotes individual autonomy will contribute to the development of a patient-
443 centered PTB registry. This patient-centered PTB registry can advance care for pregnant
444 individuals and improve birth outcomes.

445

446 *Acknowledgments:*

447 We would like to thank our study participants for their invaluable contributions.

448 **References:**

- 449 1. CDC. Preterm Birth | Maternal and Infant Health | Reproductive Health | CDC [Internet].
450 2020 [cited 2020 Dec 6]. Available from:
451 <https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pretermbirth.htm>
- 452 2. Martin JA. Births in the United States, 2019. 2020;(387):8.
- 453 3. Manuck TA. Racial and ethnic differences in preterm birth: A complex, multifactorial
454 problem. *Semin Perinatol*. 2017 Dec 1;41(8):511–8.
- 455 4. World Health Organization. World Conference on Social Determinants of Health [Internet].
456 2011 [cited 2021 Dec 12]. Available from: [https://www.who.int/news-](https://www.who.int/news-room/events/detail/2011/10/19/default-calendar/world-conference-on-social-determinants-of-health)
457 [room/events/detail/2011/10/19/default-calendar/world-conference-on-social-determinants-](https://www.who.int/news-room/events/detail/2011/10/19/default-calendar/world-conference-on-social-determinants-of-health)
458 [of-health](https://www.who.int/news-room/events/detail/2011/10/19/default-calendar/world-conference-on-social-determinants-of-health)
- 459 5. Blumenshine P, Egarter S, Barclay CJ, Cubbin C, Braveman PA. Socioeconomic
460 Disparities in Adverse Birth Outcomes: A Systematic Review. *Am J Prev Med*. 2010 Sep
461 1;39(3):263–72.
- 462 6. Shapiro-Mendoza CK, Barfield WD, Henderson Z, James A, Howse JL, Iskander J, et al.
463 CDC Grand Rounds:: Public Health Strategies to Prevent Preterm Birth. *Morb Mortal Wkly*
464 *Rep*. 2016;65(32):826–30.
- 465 7. Bublitz MH, Carpenter M, Bourjeily G. Preterm birth disparities between states in the
466 United States: an opportunity for public health interventions. *J Psychosom Obstet Gynecol*.
467 2020 Jan 2;41(1):38–46.
- 468 8. Hong X, Bartell TR, Wang X. Gaining a deeper understanding of social determinants of
469 preterm birth by integrating multi-omics data. *Pediatr Res*. 2020 Nov 13;1–8.
- 470 9. AMA. 5 things to know about clinical data registries [Internet]. American Medical
471 Association. 2014 [cited 2020 Dec 6]. Available from: [https://www.ama-assn.org/practice-](https://www.ama-assn.org/practice-management/digital/5-things-know-about-clinical-data-registries)
472 [management/digital/5-things-know-about-clinical-data-registries](https://www.ama-assn.org/practice-management/digital/5-things-know-about-clinical-data-registries)
- 473 10. Gliklich R, Dreyer N, Leavy M. Registries for Evaluating Patient Outcomes: A User’s
474 Guide. Third. Agency for Healthcare Research and Quality; 2014. 396 p.
- 475 11. Nelson EC, Dixon-Woods M, Batalden PB, Homa K, Citters ADV, Morgan TS, et al.
476 Patient focused registries can improve health, care, and science. *BMJ*. 2016 Jul
477 1;354:i3319.
- 478 12. March Of Dimes [Internet]. [cited 2020 Dec 6]. Available from:
479 <https://pretermbirthdb.org/mod/studydata>
- 480 13. Adler NE, Stead WW. Patients in Context — EHR Capture of Social and Behavioral
481 Determinants of Health. *Obstet Gynecol Surv*. 2015 Jun;70(6):388–90.

- 482 14. Buitron de la Vega P, Losi S, Sprague Martinez L, Bovell-Ammon A, Garg A, James T, et
483 al. Implementing an EHR-based Screening and Referral System to Address Social
484 Determinants of Health in Primary Care. *Med Care*. 2019 Jun;57:S133.
- 485 15. Gold R, Cottrell E, Bunce A, Middendorf M, Hollombe C, Cowburn S, et al. Developing
486 electronic health record (EHR) strategies related to health center patients' social
487 determinants of health. *J Am Board Fam Med JABFM*. 2017;30(4):428–47.
- 488 16. Zulman DM, Maciejewski ML, Grubber JM, Weidenbacher HJ, Blalock DV, Zullig LL, et
489 al. Patient-Reported Social and Behavioral Determinants of Health and Estimated Risk of
490 Hospitalization in High-Risk Veterans Affairs Patients. *JAMA Netw Open*. 2020 Oct
491 20;3(10):e2021457.
- 492 17. Goldenberg RL, Gravett MG, Iams J, Papageorghiou AT, Waller SA, Kramer M, et al. The
493 preterm birth syndrome: issues to consider in creating a classification system. *Am J Obstet
494 Gynecol*. 2012 Feb 1;206(2):113–8.
- 495 18. Kramer MS, Papageorghiou A, Culhane J, Bhutta Z, Goldenberg RL, Gravett M, et al.
496 Challenges in defining and classifying the preterm birth syndrome. *Am J Obstet Gynecol*.
497 2012 Feb 1;206(2):108–12.
- 498 19. Stout MJ, Macones GA, Tuuli MG. Accuracy of Birth Certificate Data for Classifying
499 Preterm Birth. *Paediatr Perinat Epidemiol*. 2017;31(3):245–9.
- 500 20. Bailit JL, Ohio Perinatal Quality Collaborative. Rates of labor induction without medical
501 indication are overestimated when derived from birth certificate data. *Am J Obstet Gynecol*.
502 2010 Sep;203(3):269.e1-3.
- 503 21. Lain SJ, Hadfield RM, Raynes-Greenow CH, Ford JB, Mealing NM, Algert CS, et al.
504 Quality of data in perinatal population health databases: a systematic review. *Med Care*.
505 2012 Apr;50(4):e7-20.
- 506 22. Martin JA, Wilson EC, Osterman MJK, Saadi EW, Sutton SR, Hamilton BE. Assessing the
507 quality of medical and health data from the 2003 birth certificate revision: results from two
508 states. *Natl Vital Stat Rep Cent Dis Control Prev Natl Cent Health Stat Natl Vital Stat Syst*.
509 2013 Jul 22;62(2):1–19.
- 510 23. Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health. 1991;
- 511 24. California Department of Public Health. MIHA [Internet]. [cited 2022 Mar 4]. Available
512 from: <https://www.cdph.ca.gov/Programs/CFH/DMCAH/MIHA/Pages/Default.aspx>
- 513 25. CDC. PRAMS [Internet]. 2021 [cited 2022 Mar 4]. Available from:
514 <https://www.cdc.gov/prams/questionnaire.htm>
- 515 26. County of Los Angeles Public Health. LAMB Project [Internet]. [cited 2022 Mar 4].
516 Available from: <http://publichealth.lacounty.gov/mch/lamb/lambsurvey.html>

- 517 27. LA HOPE. MCAH [Internet]. [cited 2022 Mar 4]. Available from:
518 http://publichealth.lacounty.gov/mch/LAHOPE/Provider/LAHOPE_Survey_prov.htm
- 519 28. NCFRP. FIMR [Internet]. The National Center for Fatality Review and Prevention. [cited
520 2022 Mar 4]. Available from: <https://ncfrp.org/fimr/fimr-process/>
- 521 29. UCLA. California Health Interview Survey [Internet]. [cited 2022 Mar 4]. Available from:
522 <http://healthpolicy.ucla.edu/chis/Pages/default.aspx>
- 523 30. UCSF. SOLARS [Internet]. n.d. [cited 2022 Mar 4]. Available from: <https://solars.ucsf.edu/>
- 524 31. ACOG. Prediction and Prevention of Spontaneous Preterm Birth: ACOG Practice Bulletin,
525 Number 234. *Obstet Gynecol*. 2021 Aug 1;138(2):e65–90.
- 526 32. Auger N, Le TUN, Park AL, Luo Z-C. Association between maternal comorbidity and
527 preterm birth by severity and clinical subtype: retrospective cohort study. *BMC Pregnancy*
528 *Childbirth*. 2011 Oct 4;11:67.
- 529 33. Ferrero DM, Larson J, Jacobsson B, Di Renzo GC, Norman JE, Martin JN, et al. Cross-
530 Country Individual Participant Analysis of 4.1 Million Singleton Births in 5 Countries with
531 Very High Human Development Index Confirms Known Associations but Provides No
532 Biologic Explanation for 2/3 of All Preterm Births. *PloS One*. 2016;11(9):e0162506.
- 533 34. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm
534 birth. *Lancet Lond Engl*. 2008;371(9606):75–84.
- 535 35. Reno R, Burch J, Stookey J, Jackson R, Joudeh L, Guendelman S. Preterm birth and social
536 support services for prenatal depression and social determinants. *PLoS ONE*. 2021 Aug
537 13;16(8):e0255810.
- 538 36. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall. 1991.
539 CRC press; 1990.
- 540 37. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam*
541 *Med*. 2005 May;37(5):360–3.
- 542 38. StataCorp. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP; 2015.
- 543 39. Jefferson AL, Lambe S, Chaisson C, Palmisano J, Horvath KJ, Karlawish J. Clinical
544 Research Participation among Aging Adults Enrolled in an Alzheimer’s Disease Center
545 Research Registry. *J Alzheimers Dis*. 2011 Jan 1;23(3):443–52.
- 546 40. McNeil JJ, Evans SM, Johnson NP, Cameron PA. Clinical-quality registries: their role in
547 quality improvement. *Med J Aust*. 2010 Mar;192(5):244–5.
- 548 41. Armstrong K, McMurphy S, Dean LT, Micco E, Putt M, Halbert CH, et al. Differences in
549 the Patterns of Health Care System Distrust Between Blacks and Whites. *J Gen Intern Med*.
550 2008 Jun;23(6):827–33.

- 551 42. Brandon DT, Isaac LA, LaVeist TA. The legacy of Tuskegee and trust in medical care: is
552 Tuskegee responsible for race differences in mistrust of medical care? *J Natl Med Assoc.*
553 2005 Jul;97(7):951–6.
- 554 43. Smirnoff M, Wilets I, Ragin DF, Adams R, Holohan J, Rhodes R, et al. A paradigm for
555 understanding trust and mistrust in medical research: The Community VOICES study.
556 *AJOB Empir Bioeth.* 2018 Mar;9(1):39–47.
- 557 44. DiGiuseppe DL, Aron DC, Ranbom L, Harper DL, Rosenthal GE. Reliability of birth
558 certificate data: a multi-hospital comparison to medical records information. *Matern Child*
559 *Health J.* 2002 Sep;6(3):169–79.
- 560 45. Villar J, Papageorghiou AT, Knight HE, Gravett MG, Iams J, Waller SA, et al. The preterm
561 birth syndrome: a prototype phenotypic classification. *Am J Obstet Gynecol.* 2012
562 Feb;206(2):119–23.
- 563 46. Clayton HB, Sappenfield WM, Gulitz E, Mahan CS, Petersen DJ, Stanley KM, et al. The
564 Florida Investigation of Primary Late Preterm and Cesarean Delivery: The accuracy of the
565 birth certificate and hospital discharge records. *Matern Child Health J.* 2013 Jul
566 1;17(5):869–78.
- 567 47. Gregory ECW, Martin JA, Argov EL, Osterman MJK. Assessing the Quality of Medical
568 and Health Data From the 2003 Birth Certificate Revision: Results From New York City.
569 *Natl Vital Stat Rep Cent Dis Control Prev Natl Cent Health Stat Natl Vital Stat Syst.* 2019
570 Jun;68(8):1–20.

571 Supporting information captions:

572 S1 Appendix: Domains of social determinants of health and prevalence among study groups

573 S1 File: Survey data collection instrument in English and Spanish

574 S2 File: Birth certificate data collection instrument

575 S3 File: Medical record data collection instrument

576

Fig 1. Enrollment and study groups

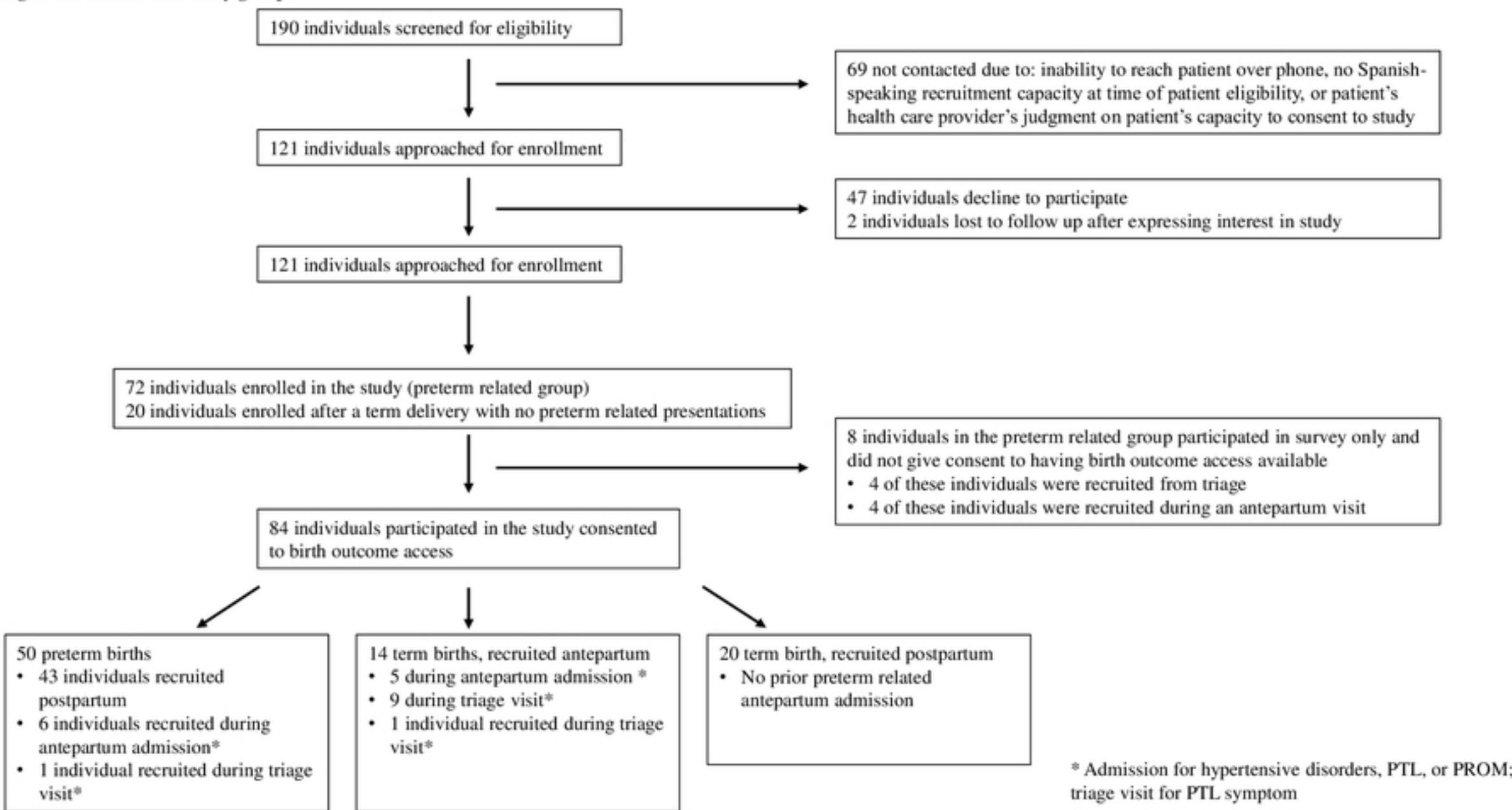
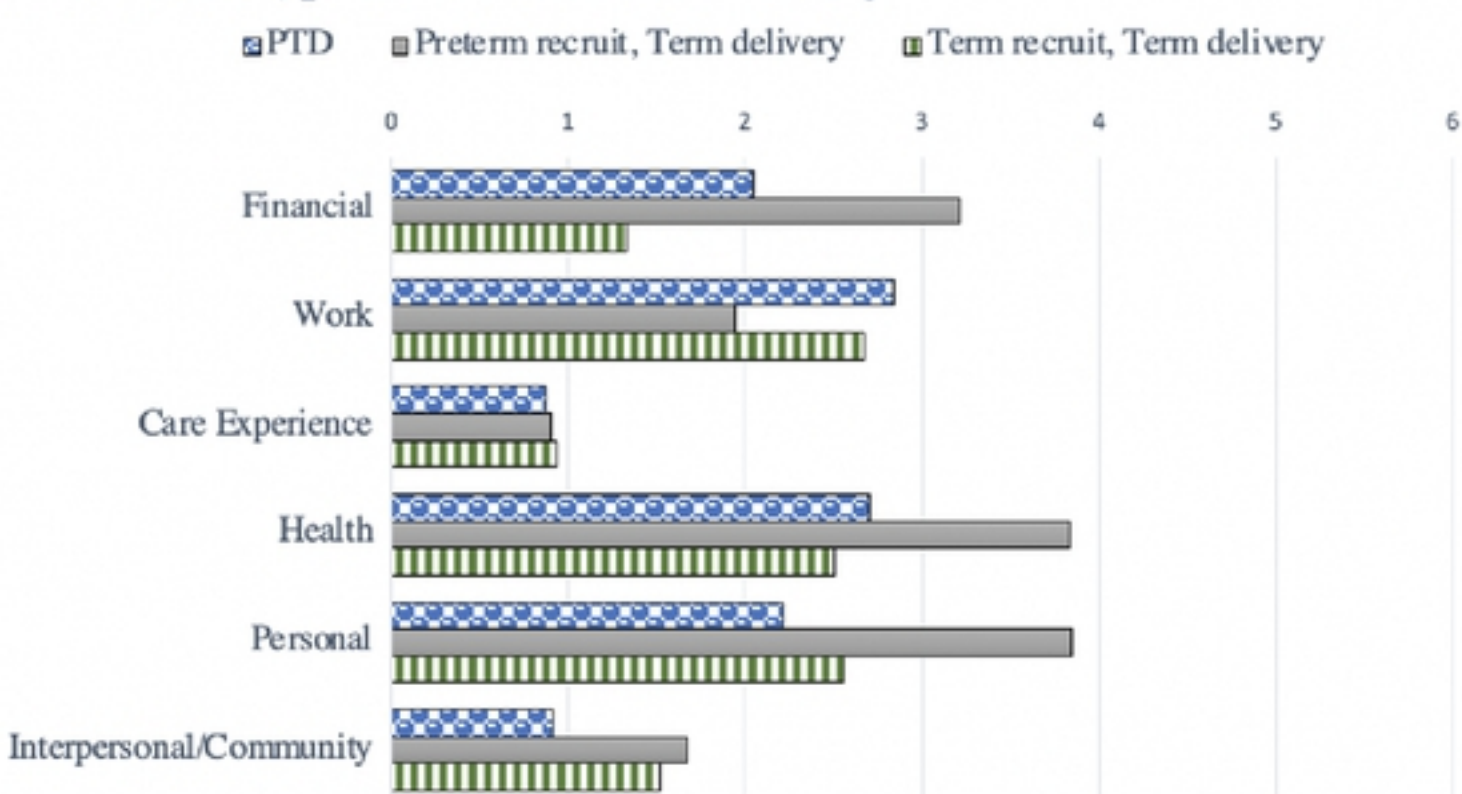


Fig 2. Structural determinants of health distributions

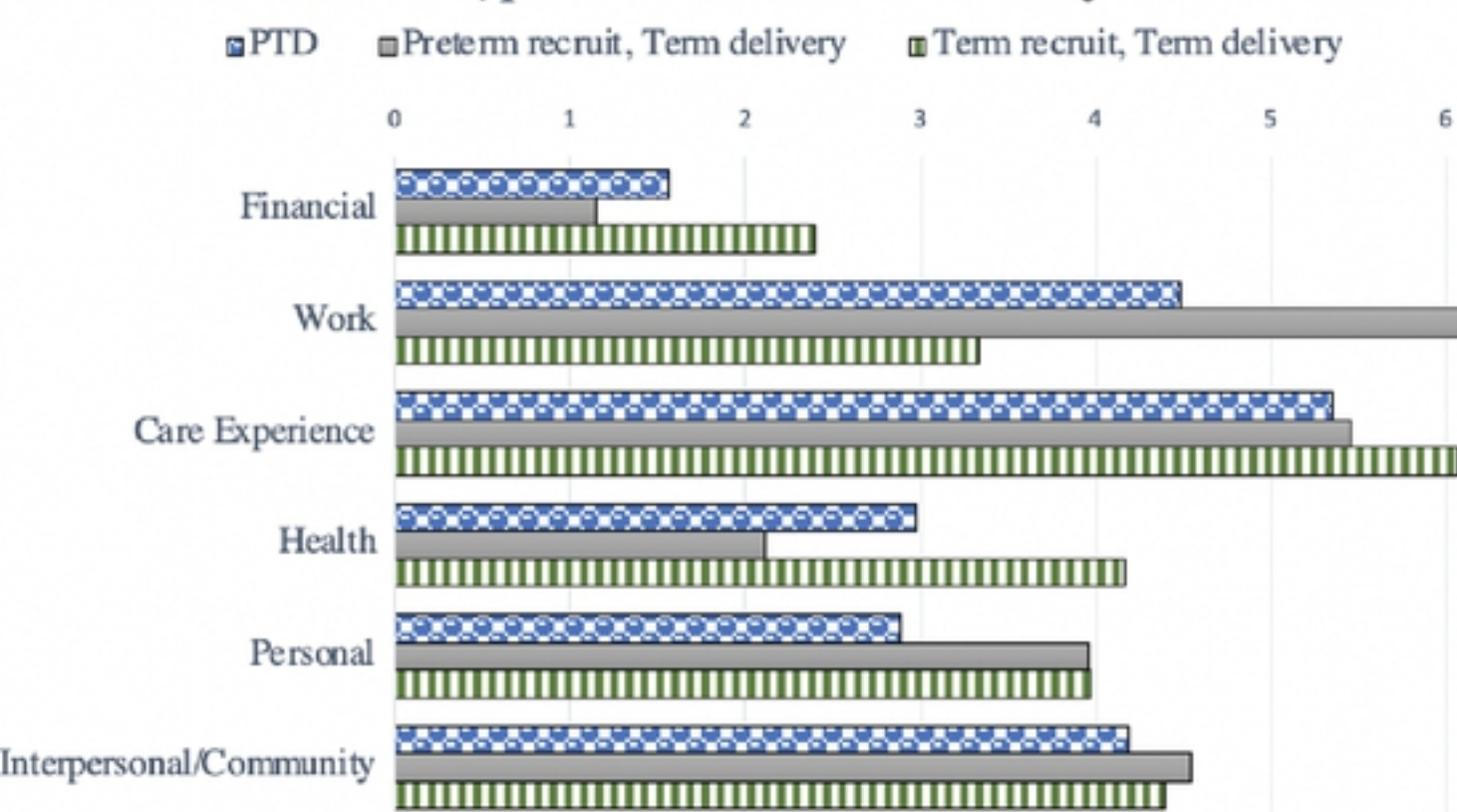
A. Stressors*, preterm versus term delivery



* Mean number of stressors in each domain normalized to scale of 0 to 10

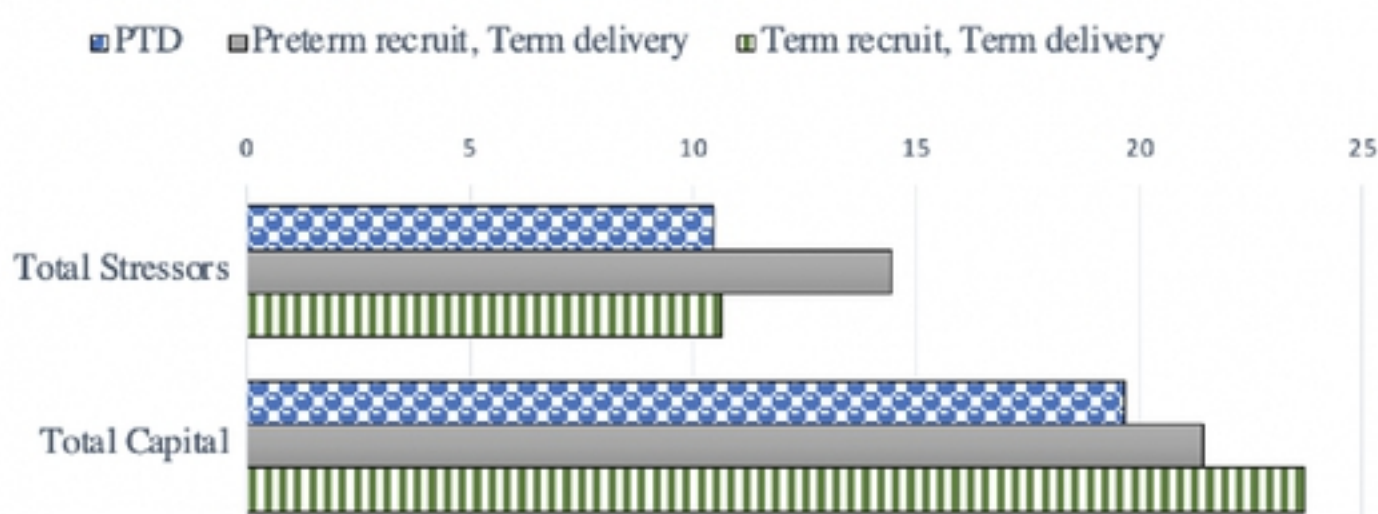
medRxiv preprint doi: <https://doi.org/10.1101/2022.03.21.22272683>; this version posted March 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

B. Protective factors*, preterm versus term delivery



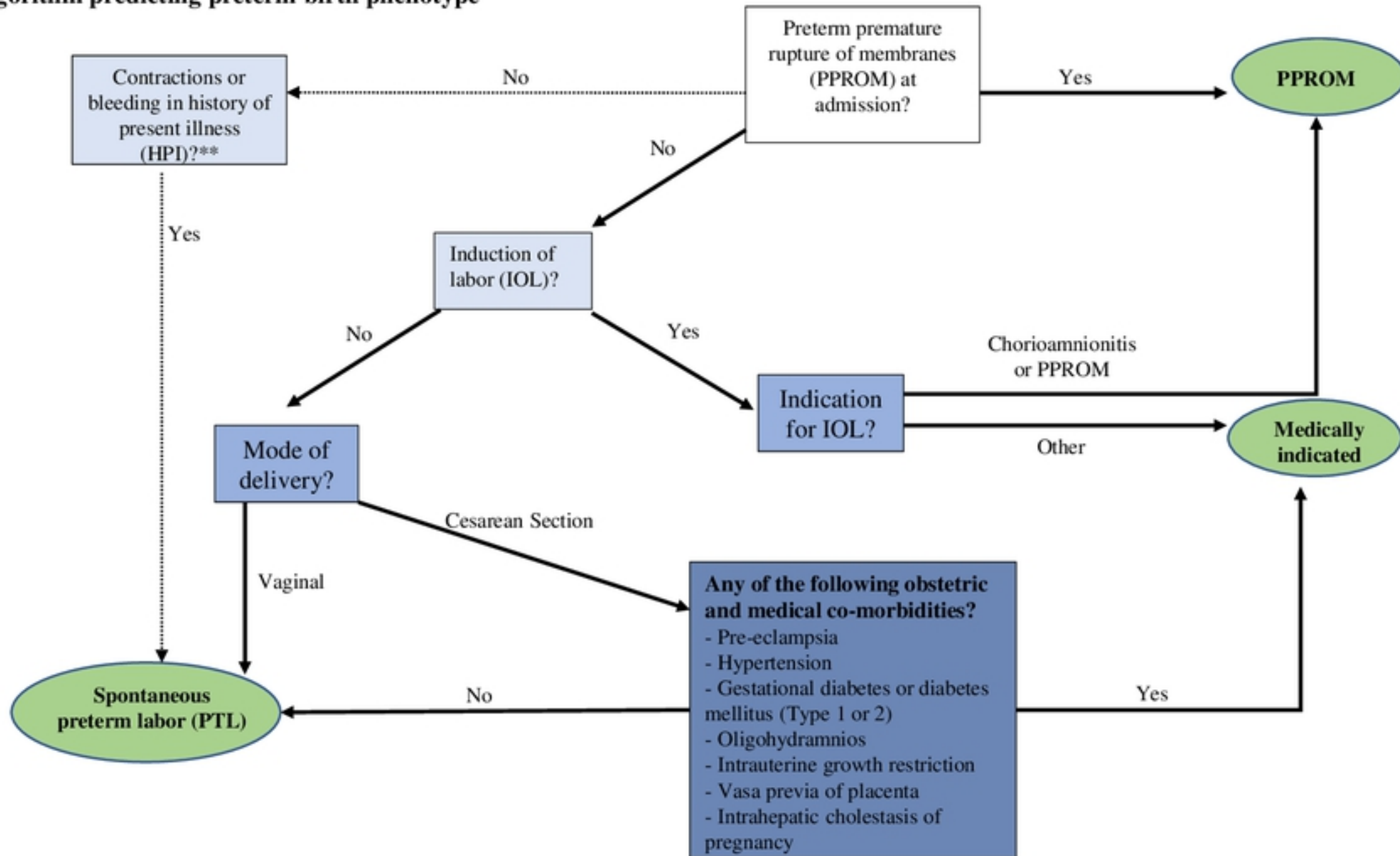
* Mean number of protective factors in each domain normalized to scale of 0 to 10

C. Total Stress and protective factors*, preterm versus term



* Sum of 6 domains each scaled 0-10.

Fig 3. Algorithm predicting preterm birth phenotype*



*Preterm birth phenotype is the final outcome indicated by the green and round box
 ***Algorithm is more accurate if signs of labor are added after the PPRM question and before IOL question

Figure