BMJ Open

Sangre Por Salud (SPS) Biobank: cohort profile

Emmil Zarrugh,[1] Marlene E Girardo,[2] Euijung Ryu,[3] Suzette Bielinski ,[3] Giovanna Moreno Garzon,[1] Valentina Hernandez,[4] Davinder Singh,[4] Sharon Levey,[1] Crystal Gonzalez,[4] Daniela B Pompa,[1] Karen Meagher,[5] Richard R Sharp,[5] Gabriel Q Shaibi,[6] Lawrence Mandarino,[7] Steven Thibodeau,[8] James R Cerhan,[2] Janet E Olson,[2] Elena De Filippis [9]

To cite: Zarrugh E, Girardo ME, Ryu E, *et al*. Sangre Por Salud (SPS) Biobank: cohort profile. *BMJ Open* 2025;**15**:e093445. doi:10.1136/bmjopen-2024-093445

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-093445).

Received 07 September 2024
Accepted 23 January 2025

For numbered affiliations see end of article.

Correspondence to
Dr Elena De Filippis;
defilippis.elena@mayo.edu

ABSTRACT

Purpose The Sangre Por Salud (SPS) Biobank was established to facilitate biomedical research opportunities for the Latino community by creating an easily accessible prospective cohort for scientists interested in studying health conditions and health disparities in this population.

Participants Individuals self-identifying as Latino, aged 18–85 years, were prospectively recruited from the primary care Internal Medicine clinic at Mountain Park Health Center in Phoenix, Arizona. After obtaining informed consent, detailed medical history questionnaires were captured, and blood samples were obtained for common laboratory tests. Participants authorised the research team to access their electronic health records for research purposes. In addition, participants had serum, plasma and DNA samples isolated and stored at the Mayo Clinic Arizona Biorepository Laboratory for long-term storage and future access. As part of the study, participants consented and agreed to be contacted for potential participation in future research studies.

Findings to date 3756 participants provided informed consent, of whom 3733 completed all study questionnaires, an oral glucose tolerance test and had blood collected and stored. The SPS cohort is predominantly composed of females (72%), with a median age at time of consent of 42 years. All participants self-identified as Hispanic/Latino, 45% were married, 53% were employed for wages and 60% had less than a high school degree. Around 25% of participants met diagnostic criteria for overweight (BMI 25–29 kg/m^2), and 49% met for obesity (BMI≥30 kg/m^2). At time of recruitment, hypertension, hyperlipidaemia and depression affected 22%, 20% and 13% of the cohort, respectively.

Future plans We plan to regularly update the participants' electronic health records and self-reported health data to longitudinal research. Additionally, we plan to obtain a more comprehensive genomic analysis on the entire cohort, ensuring greater research interest and investigation into the underlying genetic factors that contribute to disease susceptibility in this cohort.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The Sangre Por Salud (SPS) Biobank is a collaborative research initiative with Mountain Park Health Center, Arizona State University and Mayo Clinic, harnessing diverse expertise to promote research questions focused on Latino health challenges.
⇒ It exclusively recruits self-identified Latino individuals from a community primary care centre, avoiding 'helicopter research' and enhancing relevance to this population.
⇒ Participants consent to be recontacted, enabling cost-effective future studies and serving as a validation or control cohort for research comparisons.
⇒ The predominance of first-generation Americans and non-Americans among SPS participants presents challenges for studies focusing on multigenerational assessments.
⇒ The limitation of the electronic health record system, active for just over a decade, restricts access to data preceding participant consent to the biobank.

INTRODUCTION

According the 2020 Census data, the Latino population has grown by 70% since 2000, making Latino's the second-fastest growing racial and ethnic group in the USA.[1 2] Despite this demographic shift, Latinos continue to be underrepresented in biomedical and clinical research.[3–5] This lack of inclusion hinders our understanding of chronic diseases that disproportionately affect this population. According to the Centers for Disease Control and Prevention (CDC), diabetes, cancer (specifically affecting stomach, liver and women's reproductive system), heart disease and Alzheimer's disease have higher mortality rates in this population. Without a comprehensive understanding of the pathophysiological mechanisms that underlie these diseases, it is difficult to develop impactful or tailored healthcare policies, interventions, care plans and health outcome assessments that meet the unique needs of Latinos.[6] By increasing representation and enrolment of Latinos in clinical research programmes, we can begin to address these challenges and promote better health outcomes for this large portion of the US population.[6 7]

Health disparities among the Latino community are influenced by various social, behavioural and biological factors.[8–11] Understanding and addressing health disparities within the Latino community is critical to achieve health equity.[8] In parallel, increasing the participation and engagement of Latinos in biomedical research initiatives can advance scientific discoveries and care options relevant to this community.[12] Increasing representation of geographically diverse Latinos in research, based on cultural variation, regional differences, socioeconomic factors and language dialects, allows for highlighting the rich tapestry of backgrounds, experiences and cultures that contribute to the broader Latino identity while emphasising the differences shaped by location and environment. Increased participation in research will lead to a better understanding of this community's unique perspectives and experiences as they relate to their health and challenges. In contemporary healthcare, the approach of 'one-size-fits-all' for patient care is being replaced by a more personalised and individualised care approach.[13] To achieve personalised medicine, it is crucial to understand how an individual's genetic makeup interacts with their living environment and lifestyle. The understanding of these interactions is the foundation of personalised or precision medicine; where medical decisions, practices and interventions are tailored to the individual patient based on their disease risk and expected response. Efforts in personalised medicine drove the creation and utilisation of large-scale biobanks, where genetic and environmental data were collected in patients, predominantly of European ancestry, proving the successful path and precedent that can be followed when using biobanks for research purposes.[13]

More ancestrally diverse cohorts help examine current inferences about gene frequency, penetrance and pathogenicity, rather than basing these estimates on the limited biorepository data primarily gathered from population with European ancestry. However, it is also crucial to such science to recognise how many US Latino populations share genetics with American Indian, African, European and Asian populations.[14–16] In accordance with recent the National Academies of Sciences, Engineering and Medicine (NASEM, 2023) recommendations,[17] it is important to note that Latino ancestry and ethnicity can be inappropriately conflated. Genetic ancestry diversity is important for large genetic studies to capture, given how genetic variance can differ among populations that might have been geographically separate in centuries or millennia past. As the NASEM report[17] recommendations make clear, increasing and accurately describing ancestral genetic diversity in research can improve understanding of similarities (ie, our shared human heritage), rather than reifying the belief that race is a biological construct that establishes differences. Meanwhile, attending to Latino cultures and ethnicities is important to help researchers investigate gene–environment interactions and govern biorepositories in accordance with community values. Seeking more participatory models of research and health service delivery initiatives that respond to the richness of the many different traditions, which are sometimes inappropriately grouped under a single ethnic umbrella, is integral in the research of these gene–environment interactions. For example, the Sangre Por Salud (SPS) biobank was created with the active participation and involvement of Mexican and Mexican American community members in Phoenix, Arizona. Other Latino biobanks might yield distinctive scientific insights regarding the causal interplay among biological factors, environmental exposures, and social determinants of health. This could be, in part, related to a diverse representation of Hispanic and/or Latino groups. Ultimately, integrated analyses of data from all major Latino cohorts could provide a comprehensive understanding of various health-related challenges faced by this community.

In 2013, the Mayo Clinic Center for Individualized Medicine established a Biorepository in collaboration with the Mountain Park Health Center and Arizona State University (ASU) in the Phoenix metropolitan area of Arizona.[18] The biorepository aimed to increase racial and ethnic diversity in the Mayo Clinic Biorepositories Program[19] and enable access to precision medicine for a broader range of patients, including a representation of the Latino. In creating this biorepository, we used the opportunity to advance biomedical discoveries and translate them into practical diagnostic and therapeutic options tailored specifically for the community. Also, we intended to implement these novel options in the participants' community clinical setting, when feasible, to enhance healthcare accessibility and educate Latinos on their health to promote better health outcomes. The Mayo Clinic Biobank protocol, established in 2009 in several locations across the USA, including Rochester Minnesota, Jacksonville Florida and La Crosse and Onalaska Wisconsin, features over 56 000 participants, with 90% self-identifying as white.[19] This has enabled close to 300 approved research projects to access the biobank data and/or samples, with the distribution of over 200 000 sample aliquots for research purposes. Due to the existence of the Mayo Clinic Biobank and the precedent it has established, the need to create similar biobanks for under-represented communities, such as the Latino population, became evident. To address this need, SPS Biobank, which translates as 'Blood for Health,' was established to promote biomedical research opportunities for Latinos in the US. From 2013 to 2018, the SPS Biobank successfully enrolled 3756 participants who self-identified as Latinos. The primary goal of SPS is to facilitate collaborative and multidisciplinary research projects that aim to understand and improve the health of the Latino population. By collecting and analysing biological samples and health data from participants, SPS provides a valuable resource for researchers seeking to address health disparities and improve healthcare outcomes for Latinos in this region of the USA.

COHORT DESCRIPTION

Study design

The recruitment model for SPS Biobank has been described.[18] Briefly, to expand precision medicine research to an under-represented Latino patient population, the Mayo Clinic Center for Individualized Medicine leveraged the previous experience and infrastructure developed in collaboration with Mountain Park Health Center, a Federally Qualified Health Center that provides care to a large community of Latinos in the Phoenix metropolitan area and ASU. A certificate of confidentiality from the National Institutes of Health was obtained to protect the privacy of human research participants enrolled in the study.

Participant enrolment

Participant enrolment for SPS Biobank began in May 2013, when providers within the primary care Internal Medicine Department at Mountain Park Health Center approached eligible participants during medical encounters. Participants were self-identified Latinos between 18 and 85 years. Patients with a medical history of cancer with evidence of remission for more than 3 years were included in the recruitment. Exclusion criteria included active cancer under treatment (excluding non-melanoma skin cancers) within the past 3 years, current pregnancy, postpartum within the last 12 months or currently breastfeeding. Individuals who were unable to abstain from smoking for 4 hours were also excluded. From 2013 to October 2015, we also excluded patients with diabetes mellitus. However, due to the high prevalence of type 2 diabetes among Latinos, we reviewed our exclusion criteria and agree to include patients with type 2 diabetes to enhance the phenotypic diversity of the SPS Biobank and to make this cohort more representative of the Latino community in Arizona. Interested participants were offered an overview of the SPS Biobank during their medical encounter, and those who expressed further interest were referred to an SPS Biobank study team member.

Data collection

To collect data for the SPS Biobank, interested research volunteers, who were found eligible, were scheduled for a morning study visit after an overnight fast. On providing informed consent, participants completed a comprehensive questionnaire that collected data related to participants' medical, family, social history and lifestyle information including exercise and dietary habits. As part of the demographic information, we collected self-identified race (white, black or African American, Asian, Native Hawaiian or other Pacific Islander, American Indian or Alaska Native, other to be specified) and ethnicity (Latino or non-Latino), as defined by the US Census Bureau 2013. Anthropometric parameters, including temperature (to rule out acute illnesses), seated blood pressure and pulse, as well as height, weight and waist circumference (measured in triplicate and averaged) were then obtained. All data collected were stored in a password-protected Research Electronic Data Capture (REDCap) system at the Mayo Clinic, accessible only to IRB-approved Mayo Clinic research personnel.

Biospecimen samples

Participants were asked to provide fasting blood samples for various assessments. The collection process involved obtaining two 10 mL tubes with EDTA to evaluate clinical laboratory parameters, one 10 mL tube with no additives to isolate and store serum, one 4.5 mL tube with sodium citrate for coagulation studies and an additional 1 mL EDTA tube for buffy coat and plasma.

Measures of clinical laboratory parameters, such as a complete blood count with differentials, lipid profiles, haemoglobin A1C (HbA1c) and a basic metabolic panel, were completed for all participants. The clinical laboratory analyses were collected and run by Sonora Quest Laboratories, a commercial laboratory licensed and certified under the Clinical Laboratory Improvement Amendments of 1988. The clinical test cost was covered by Mayo Clinic, but all results were reported back into the patient's electronical health record at Mountain Park Health Center. This would provide an immediate benefit for those patients who were uninsured or had economic challenges, by relieving them from the burden of high cost for laboratory analyses needed to outline their plan of care. The primary care providers in the Internal Medicine clinic at Mountain Park Health Center received a copy of all patients' results, deposited in their health record system. For those with clinically significant results, providers followed up and managed as appropriate.

All participants underwent a 75 g oral glucose tolerance test (OGTT) in accordance with the criteria established by the American Diabetes Association to confirm normal glucose tolerance.[20] Those who met the diagnostic criteria for diabetes mellitus were referred to their Internal Medicine clinicians for additional care. Patients with a previous diagnosis of diabetes mellitus did not complete the OGTT procedure.

From the buffy coat, white blood cells were isolated for DNA extraction, while plasma was stored for future studies. All remaining samples were processed, aliquoted and stored at −80°C in the Biospecimens Accessioning and Processing Laboratory at the Mayo Clinic in Scottsdale, AZ. Along with stored samples, the SPS Biobank has also completed genome-wide genotyping on the entire cohort to provide access to genomic data.

Patient and public involvement

Community advisory board

In addition to recruiting research volunteers to participate in our biorepository, we involved members of the Latino community to participate and be the voice of the local population. We established a bilingual/bicultural Community Advisory Board (CAB) within the Latino community that would collaborate with researchers to ensure overall transparency, to involve members of the

Latino community in discussions of complex ethical issues, and to provide guidance for the overall governance of the SPS biobank. The CAB opinion is obtained in different ways. Some examples are here reported.

The study team first requested CAB input to identify effective strategies to help with increasing participant enrolment in the study. The study team held many discussions with the CAB around study start-up decisions, including a discussion on whether inability to read or write or both in their native language should be considered an exclusion criterion. The CAB strongly advised against having such inability as an exclusion criterion and the study team accepted this executive decision. The CAB advised instead designing recruitment and consent procedures that supported the decision-making agency of all Mountain Park Health Center patients eligible to express a voluntary and informed decision to participate in the biobank. The CAB was also involved in discussions related to industry access to the SPS cohort for research purposes. Currently, the CAB has mostly been involved in discussions around potential genomic analysis and the development of a return in results procedures for actionable and non-actionable genetic results.

Multi-institutional collaboration involves ongoing engagement, troubleshooting and relationship-building. Some of our lessons learnt might be helpful to those seeking to currently increase their collections' diversity. To date, the SPS Biobank engagement efforts include:

- Disseminating SPS-related content and research updates via bilingual newsletters.
- Sharing research-related clinically actionable results with Mountain Park Health Center clinicians in internal medicine meeting updates.
- Encouraging SPS utilisers to engage directly with the CAB, integrating community attitudes and perspectives at various stages of a research project including study proposal, external partners, study design, aggregate or individual results, or research implications.

The details on the CAB organisation, contributions and roles were previously described.[18]

Electronic health record data

Participants agreed to share with the research team their health record data starting from 1 January 2013, or from their first enrolment visit at Mountain Park Health Center, whichever came first. The update is completed every 6 months. We collect data related to the following categories demographics, anthropometrics, laboratory analysis, International Classification of Diseases (ICD) codes, Current Procedural terminology codes, medications and social determinants of health. The opportunity to collect electronic health record (EHR) data on an ongoing basis allows the research team to track their health outcomes.

Statistical analyses

Enrolment data for age at consent, sex and measured in clinic height and weight to calculate body mass index (BMI) were collected, while self-reported questionnaires were used to obtain information about insurance, education, smoking and alcohol status. The characteristics of SPS Biobank participants were summarised as percentages and counts. To contextualise these findings, we compared the characteristics of SPS participants to those of Hispanic/Latino participants in the CDC's 2019 Behavioral Risk Factor Surveillance System Survey (BRFSS: https://www.cdc.gov/brfss/), which collects behavioural health risk data through telephone interviews within each state. Both the Arizona total and Southwest (Arizona, California, Colorado, Nevada, New Mexico and Utah) BRFSS surveys were used for comparison. Additionally, we summarised the top 10 self-reported health conditions and 15 of the top 20 chronic conditions defined by the US Department of Health and Human Services.[21] The five conditions excluded (ie, schizophrenia, autism, HIV, dementia and cancer) had a prevalence of <0.2%. The prevalence of each chronic condition was defined as the proportion of participants with an ICD 9/10 code and the date of the code prior to the consent date. After excluding prevalent cases, the incidence rate of each condition until the last follow-up date was calculated using Kaplan-Meier method. The summary statistics of laboratory measurements obtained at enrolment were presented using the median, range and IQR. Diabetic status was determined based on baseline laboratory measurements including fasting glucose, HbA1c and 2-hour post OGTT glucose. The laboratory measurements were categorised into normal ranges before generating counts and percentages.

FINDING TO DATE

Recruitment

10 813 patients were first identified as potential candidates for the SPS Biobank based on their EHRs. However, 5010 patients (46%) declined participation, because they were not interested, or during interview with the research coordinator were found to meet some exclusion criteria previously missed. In some cases, they were not able to be reached by phone. When they were reached, the main reasons for declining participation were lack of transportation to participate in person, scheduling conflicts, time constraints or unspecified reasons. Out of the remaining patients, 5803 agreed to participate and were scheduled for a research appointment. Unfortunately, 2047 participants (35%) did not attend their scheduled appointment or were no longer interested in participating. Ultimately, after reviewing the study details, 3756 participants provided informed consent. Of those, 3733 completed all study procedures after providing informed consent (online supplemental figure 1).

On the study's initiation, achieving the anticipated recruitment goals proved more challenging than predicted. For instance, during the first year, we approached 595 eligible participants, but 199 declined to participate for the reasons mentioned above. Of the remaining 399, only 328 successfully completed all the study procedures. Weekly, we began collecting data on

Table 1 SPS recruitment metrics over the course of the enrolment period 2013–2018

Year	2013	2014	2015	2016	2017	2018	Total
Approached	595	1054	1650	2670	3593	1251	10813
Consented	328 (55%)	676 (64%)	743 (45%)	885 (33%)	899 (25%)	225 (18%)	3756 (35%)

SPS, Sangre Por Salud.

the number of subjects approached who declined versus scheduled versus completed study visits. Through this process, we identified multiple gaps within our recruitment strategy, including the lack of understanding of the study goals, low literacy level and concerns related to the US political climate and immigration status. Additionally, our limited lack of sufficient resources further hindered our overall recruitment effort and goals. To improve our accrual, we cross-trained our team members to better allocate our resources on study days where they were most needed. We also established shared team goals to enhance our team morale and motivation. To accommodate transportation limitations, we took advantage of a transportation service provided by Mountain Park Health Center. Lastly, we involved the CAB and used their input to help address social barriers that were limiting our research enrolment. The combination of these strategies and efforts from the study team led to more eligible participants being approached and ultimately accrued (table 1).

Cohort summary

SPS cohort is composed of mostly females (72%), with a median age of 42 years. 25% of the participants were categorised as overweight (BMI 25–29.9 kg/m^2), and 49% had obesity as defined by their BMI≥30 kg/m^2 (table 2). The Latino population included in our biorepository exhibits high rates of metabolic syndrome, based on their BMI, as well as a self-reported or documented diagnosis of hypertension and hyperlipidaemia in their EHRs. Table 3 shows the lipid profile results collected at enrolment and the percentage of HbA1c collected during the study visit. Analysis of the laboratory data and results from the OGTT revealed many patients with undiagnosed diabetes mellitus. In our cohort, 534 individuals (14%) had a HbA1c level greater than 6.4%, 88 individuals (3%) had a fasting glucose level greater than 125 mg/dL and 191 individuals (6%) had a 2-hour post OGTT glucose level greater than 199 mg/dL (table 4).

The SPS Biobank participants have been involved in the study for an average of 3.95 years (figure 1A) and have an average of 4.39 visits per year (figure 1B). 32% of the participants have been seen at least once every year for the past 6 years (table 5). The density of different data elements in the EHR, including diagnosis codes, procedure codes, laboratory test results, vitals and prescribed medication, has been tracked from 2013 to 2020 (figure 1C).

SPS participants as representative of the Latino community of Arizona

Compared with the data collected in the 2019 CDC's BRFSS survey specific for the Latino group in Arizona and the Southwest (including Arizona, California, Colorado, Nevada, New Mexico and Utah), SPS participants were younger (57% aged 18–44 years vs 43% in BRFSS AZ and 50% in BRFSS Southwest), higher partecipation of female (72% vs 54% in BRFSS AZ), self-defined multiracial (24% vs 2% in BRFSS AZ and 1.4% in BRFSS Southwest) and had a higher prevalence of obesity (49% vs 36% in BRFSS AZ and 32% in BRFSS Southwest) (table 2). A possible explanation for such discordance could be related to the methods adopted for ascertaining racial background and BMI. In table 2, we present the self-identified racial background captured via questionnaire, however, when we extracted the same variable from the EHR, our cohort representation doubled for whites (88% by EHR vs 47% by SPS self-reported questionnaire), while the multiracial dropped to level similar to what reported in the BRFSS (0.2% by EHR vs 1.4% in the BRFSS AZ) (online supplemental table 1). For BMI, the discrepancy may be due to the difference between self-reported data (BRFSS) versus measured data. In SPS a research coordinator measured height and weight to then calculate BMI.

We also compared the sociodemographic characteristics between SPS participants and the BRFSS cohorts. We observed that BRFSS AZ reported 78% of participants to be smokers, while 68% admitted smoking in the SPS cohort. This percentage was similar when compared with the BRFSS Southwest, where 67% of participants smoked. SPS participants were more likely to report consuming alcohol at least once a month (28% in SPS vs 9% in BRFSS AZ and 10% in BRFSS Southwest). The majority (60%) of SPS Biobank participants had an eighth grade level education in a much higher proportion than the other two cohorts, which were 27% in BRFSS AZ and 30% in BRFSS Southwest, respectively.

Additionally, possibly because the study protocol covered clinical laboratory cost, the SPS cohort had a much higher percentage of uninsured participants (65%) compared with BRFSS AZ (22%) and BRFSS Southwest (22%).

Given that 72% of the SPS cohort is female, we completed a separate analysis adjusted for sex-specific rates (online supplemental tables 2,3). The results of this analysis showed differences in smoking status, alcohol consumption, which were more often reported by men participating in SPS than men in BRFSS (online

Table 2 Characteristics of SPS Biobank participants, compared with the Center for Disease Control's 2019 BRFSS Hispanic/Latino participants (Southwest regions include AZ, CA, CO, NV, NM, and UT)

	SPS Biobank (N=3733)	BRFSS Southwest (N=8298)	BRFSS AZ (N=1896)
Age at enrolment, n (%)			
18–44	2139 (57)	4125 (50)	936 (49)
45–54	914 (25)	1410 (17)	332 (18)
55–64	454 (12)	1268 (15)	310 (16)
65 or older	226 (6.1)	1495 (18)	318 (17)
Sex, female n (%)	2655 (72)	4470 (54)	1042 (55)
Race, n (%)			
White	3285 (88)	5247 (71)	1113 (67)
Black/African American	8 (0.2)	132 (1.8)	20 (1.2)
American Indian/ Alaskan Native	16 (0.4)	446 (6.0)	82 (4.9)
Asian	3 (0.1)	34 (0.5)	7 (0.4)
Native Hawaiian/ Pacific Islander	2 (0.1)	27 (0.4)	2 (0.1)
Other	393 (11)	1352 (18)	422 (25)
Multiracial	7 (0.2)	148 (2.)	24 (1.4)
Missing	19	912	226
Not insured, n (%)	2413 (65)	1828 (22)	434 (23)
Missing	1	62	18
Body mass index, n (%)			
Underweight	23 (0.6)	111 (1.5)	14 (0.9)
Normal Weight	583 (16)	1970 (27)	400 (25)
Overweight	1301 (35)	2658 (37)	583 (36)
Obesity	1798 (49)	2544 (35)	610 (38)
Missing	28	1015	289
Education, n (%)			
Less than high school	2024 (60)	1872 (23)	397 (21)
High school graduate	694 (20)	2642 (32)	596 (32)
Associate degree/ technical school	475 (14)	2100 (25)	490 (26)
College graduate or more	188 (5.6)	1651 (20)	406 (21)
Missing	352	33	7
Smoking, n (%)			
Current	267 (8.0)	947 (12)	905 (52)
Former	465 (14)	1541 (20)	160 (9.1)
Never	2612 (78)	5292 (68)	365 (19)
Missing	389	518	139
Alcohol, n (%)			
Never	1636 (48)	4236 (55)	905 (52)
Once a month	969 (28)	669 (8.6)	160 (9.1)
2–4 times a month	589 (17)	1534 (20)	365 (21)
2–5 times a week or more	229 (6.7)	1340 (17)	327 (19)
Missing	310	519	139
Employment, n (%)			
Employed for wages	1823 (53)	4589 (56)	1070 (58)
A homemaker	1046 (31)	698 (8.6)	172 (9.3)
A student	89 (2.6)	328 (4.0)	73 (4.0)
Retired	71 (2.1)	1224 (15)	253 (14)
Unable to work	157 (4.6)	565 (6.9)	124 (6.7)
Not currently employed	233 (6.8)	744 (9.1)	156 (8.4)
Missing	314	150	48
Marital status, n (%)			
Married	1679 (45)	3827 (46)	861 (46)
Divorced	220 (5.9)	970 (12)	239 (13)
Widowed	104 (2.8)	492 (6.0)	100 (5.3)
Separated	198 (5.3)	336 (4.1)	70 (3.7)
Never married	1098 (29)	2029 (25)	491 (26)
A member of an unmarried couple	425 (11)	593 (7.2)	119 (6.3)
Missing	9	51	16

AZ, Arizona; BRFSS, Behavioural Risk Factor Surveillance System Survey; CA, California; CO, Colorado; NM, New Mexico; NV, Nevada; SPS, Sangre Por Salud; UT, Utah.

supplemental tables 2,3). Employment among men was also different than the other two cohorts examined (online supplemental tables 2,3).

In our health questionnaire, we asked participants about 60 different health conditions. Table 6 displays a summary of the top 10 self-reported illnesses. The most frequently self-reported condition was hyperlipidaemia (19%), followed by depression (18%) and hypertension (18). Table 7 reports the top 15 chronic conditions with the highest prevalence in our cohort, where prevalence was defined as the proportion of patients with each condition prior to enrolment into SPS Biobank. Hypertension (22%), hyperlipidaemia (20%) and depression (13%) were the most common conditions observed in our study. The remaining conditions (schizophrenia, autism, HIV, dementia and cancer) with a prevalence of less than 0.2% were excluded from our analysis.

Cohort utilisation

Genotype-driven research investigations

Using data generated by genome-wide association study, Standage-Beier *et al*[22] evaluated the vitamin D receptor (VDR) single-nucleotide polymorphisms (SNPs) and its association with cardiometabolic phenotypic data in the SPS cohort. The authors identified that in the Latino population analysed, VDR SNPs were associated with several measures of cardiometabolic syndrome particularly with fasting glucose, HbA1c and BMI. As research investigators gain a better understanding of the role

Table 3 Laboratory measurement at the SPS Biobank consent

	Normal		Abnormal		Overall
	N (%)	Median (range)	N (%)	Median (range)	Median (range)
HbA1c (%)	1588 (43)	5.4 (4.1–5.6)	2127 (57)	6 (5.7–6.8)	5.7 (4.1–16.8)
Cholesterol (mg/dL)	2595 (70)	166 (49–199)	1129 (30)	221 (200–459)	180 (49–459)
Triglyceride (mg/dL)	2435 (65)	97 (27–149)	1289 (35)	197 (150–1876)	121 (27–1876)
HDL cholesterol (mg/dL)	2159 (58)	58 (50–223)	1565 (2)	41 (3–49)	47 (3–223)
LDL cholesterol (mg/dL)	1613 (44)	82 (20–99)	2041 (56)	123 (100–343)	104 (20–343)
Non-HDL (mg/dL)	1713 (48)	107 (29–129)	1865 (52)	156 (130–382)	131 (29–382)
VLDL cholesterol (mg/dL)	2400 (66)	19 (7–29)	1259 (34)	38 (30–80)	24 (7–80)

Normal ranges for HbA1c: ≤5.6, cholesterol: ≤199, triglyceride: ≤149, HDL cholesterol: ≥50, LDL cholesterol: ≤99, non-HDL: ≤129, VLDL cholesterol: ≤29.
N=3733.
HbA1c, haemoglobin A1C; HDL, High-Density Lipoprotein; LDL, Low-Density Lipoprotin; SPS, Sangre Por Salud; VLDL, Very Low-Density Lipoprotein.

VDR plays in metabolic diseases, this knowledge would be of great value at the time of future drug-targeted development.

More recently, Arora *et al*[23] assessed the frequency of gallstone disease (GSD) in the SPS Biobank and evaluated for potential genetic variants associated with cholelithiasis. Previously, the prevalence of GSD in the Mexican American population was reported up to 26.7% in females, and only 8.8% in males.[24] 172 incidences of GSD were identified in the SPS cohort and 87% of these cases were detected in females. Analysis of GWAS studies completed in various populations has identified several population-specific genetic variants associated with GSD.[25] In the Mexican American population, specific genetic variants within four genes were identified.[26] In our Latino cohort, the authors identified intron variants within MATN2, GPC6 and GPRIN3 genes that increase the risk for GSD. While the study results need to be confirmed in a wider sample, this analysis may have identified new biological pathways involved in the pathogenesis of GSD within a Latino population living specifically in the Southwestern region of the USA.

Precision medicine in the SPS Latino cohort

To ensure that precision medicine initiatives are widely implemented, it is crucial for large academic institutions to partner with community-based clinical practices such as MPHC. One example of such a partnership is the SPS Biobank's involvement in the Electronic Medical Records and Genomic Consortium III Network, specifically the 'Arizona Return of Actionable Variants Empirical Study.' In this study, the authors developed resources to engage clinical providers in the project and created a process for returning both actionable and non-actionable results to research participants and their providers using the EHR.[27] They also developed a process to provide clinical decision support for additional needed clinical care.

Over 12 months, 500 patients originally enrolled in SPS were recalled and agreed to be enrolled in this study. A newly IRB-approved informed consent was provided to these participants. Analysis of these data identified 10 participants with a pathogenic/likely pathogenic variant. However, the study group encountered challenges in returning results, including difficulty reaching patients, a lag time between enrolment and return of results, low health literacy, lack of health insurance and limited family history information. The authors concluded that addressing social determinants of health is a fundamental step in the success of genomic medicine initiatives conducted in underserved communities.[28]

Lifestyle and wellness factors related to cardiometabolic disease

Among various ethnic groups, Latinos have been found to have one of the highest prevalence rates for cardiometabolic syndrome. A previous study indicated that this increased risk could potentially be modified.[29] Dugani *et al* also conducted a study to identify modifiable risk factors, such as lifestyle factors and wellness measures,

Table 4 Diabetes status by different laboratory measurements at baseline (N=3733)

	HbA1c		Fasting glucose		2 hour glucose	
	Test value	N	Test value	N	Test value	N
Normal	<5.7%	1588	<100 mg/dL	2525	<140 mg/dL	2461
Pre-diabetes	5.7%–6.4%	1593	100–125 mg/dL	598	140–199 mg/dL	548
Diabetes	>6.4%	534	>125 mg/dL	88	>199 mg/dL	191

HbA1c, haemoglobin A1C.

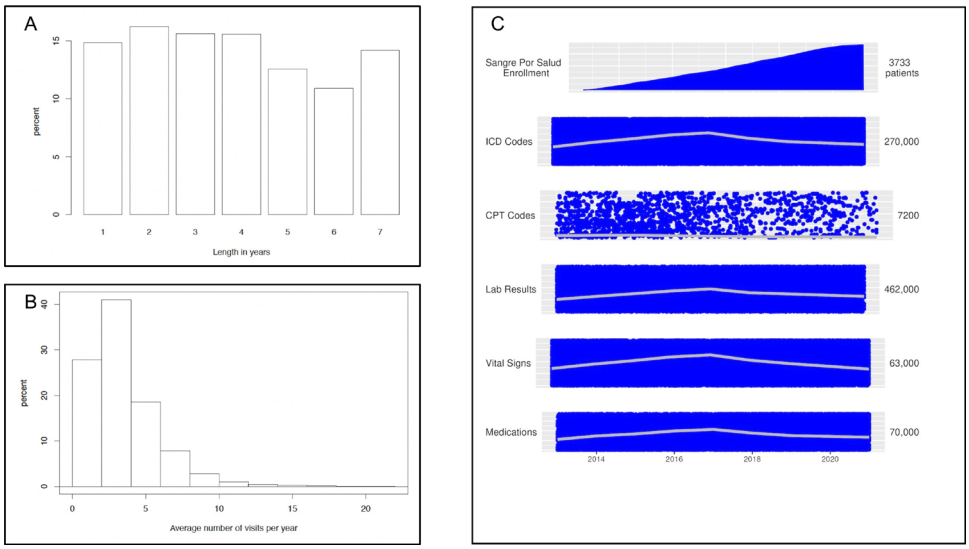

Figure 1 Length and depth of electronic health records. (A) Difference in years between the last visit date and the first visit dates. (B) Average number of visits per patient per year. (C) Density of electronic health record elements captured between 2013 and 2020, these elements include diagnosis codes, procedure codes, laboratory test results, vital signs and prescription medication. SPS participants began enrolling in 2013 and data currently extends through the end of 2020. Each blue dot indicates a single data point on a certain day for each patient while the grey line represents the proportion of participants that have a record for that each given year. CPT, current procedural terminology; ICD, International Classification of Diseases; SPS, Sangre Por Salud.

that contribute to the clinical risk of developing pre-diabetes or type 2 diabetes in Hispanic adults.[30] This analysis confirmed that modifiable cardiometabolic and lifestyle factors were associated with an increased risk of pre-diabetes and new diagnoses of type 2 diabetes in our Latino cohort, which aligns with previous reports. The results of this study could provide a basis for tailored interventions to prevent the onset of cardiometabolic syndrome among Latino populations.

Strengths and limitations

The SPS Biobank project has several strengths that set it apart from other biobanks. First, our collaboration spans three institutions: Mountain Park Health Center, ASU and Mayo Clinic. This partnership allows us to draw on diverse expertise in research, clinical care and understanding of the Latino community, all of which are crucial in facilitating and conducting high-quality research that addresses the unique needs of this population.

Unlike other biobanks that are established within the walls of academic tertiary centres, our cohort is composed solely of self-identified Latino individuals recruited from a community primary care centre. This approach enables us to avoid 'helicopter research' and partner with clinicians providing care in the community, ensuring that our research is more relevant and responsive to the needs of the Latino population. Additionally, this approach has allowed us to access EHR data and has enriched

Table 5 Depth of available electronic health record data for SPS Biobank participants (2013–2020)

	Summary (N=3733)
Average length (in years)	3.95
Average number of visits per year	4.39
Number of years with ≥1 visit, N (%)	
0	1 (0)
1	528 (14)
2	541 (15)
3	453 (12)
4	521 (14)
5	497 (13)
6	389 (10)
7	358 (9.6)
8	444 (12)

SPS, Sangre Por Salud.

Table 6 Top 10 self-reported health conditions by SPS participants on the self-assessment questionnaire

Health conditions	N	Frequency
Hyperlipidaemia	653	19%
Depression	629	18%
Hypertension	628	18%
Abnormal distance vision	548	17%
Anxiety	519	15%
Migraine	306	9.7%
Asthma	306	8.9%
Osteoarthritis	291	8.5%
Gastro-oesophageal reflux disease	257	8.2%
Type 2 diabetes	275	8.0%

SPS, Sangre Por Salud.

Table 7 Top 15 of the 20 US Department of Health and Human Services (DHHS) chronic conditions via electronic health records

Health conditions	Prevalent cases (%*)	Incident rate (per 100 person-years)
Hypertension	830 (22%)	3.8
Hyperlipidaemia	766 (21%)	18.1
Depression	482 (13%)	3.5
Diabetes	267 (7.2%)	0,8
Asthma	196 (5.3%)	1.1
Arthritis	111 (3.0%)	1.1
Substance abuse	87 (2.3%)	0.7
Chronic kidney disease	43 (1.2%)	0.4
Coronary artery disease	37 (1.0%)	0.05
Hepatitis	24 (0.6%)	0.3
Chronic obstructive pulmonary disease	21 (0.6%)	2.9
Osteoporosis	18 (0.5%)	0.03
Stroke	17 (0.5%)	0.03
Congestive heart failure	14 (0.4%)	0.2
Cardiac Arrhythmias	11 (0.3%)	0.12

*Schizophrenia, autism, HIV, dementia and cancer were excluded due to prevalence <0.2%.

our dataset with in-depth questionnaires, capturing self-reported information on medical history and lifestyle. Thus, this additional data complements the biological samples collected and makes the SPS Biobank a valuable resource for researchers interested in understanding the prevalent health challenges that face the Latino community.

Additionally, participants agreed to be recontacted for updates or to explore new research questions or interests. If successful, recontacting participants will allow for cost-effective opportunities to invite volunteers to participate in subsequent studies. Additionally, our cohort can serve as a validation cohort for comparison to findings collected in other research settings and as a control cohort matched to a disease cohort with the same ethnic or genetic background. This also serves as another cost and time-effective opportunity to be used by researchers and investigators.

The SPS Biobank's completion of a genome-wide genotyping has provided access to genomic data that could be instrumental in gaining insight into specific phenotypes observed in the cohort. However, the discovery of incidental findings of actionable genes raises ethical questions about how to inform research participants while providing the required care without overwhelming them, their family or their medical providers.

While the SPS Biobank has many strengths, it is important to acknowledge its limitations. One limitation is that the population included in the cohort is relatively young, mostly female and with few chronic diseases. This may result in an incomplete representation of health problems in the Latino community. Additionally, the significant number of young, first-generation, non-Americans in the SPS biobank may pose challenges during recontact processes and further accessibility. To address this, the SPS research team has requested that participants identify a next-of-kin who can provide updated demographic information if needed. Furthermore, the fact that most SPS Biobank participants were born outside the USA may account for some of the differences observed compared with other cohorts, such as the BRFSS AZ and Southwest cohorts. It should be said that the nature of our study where participants had to come to the clinic during regular business hours and weekdays, may have created some selection bias in the socioeconomic class represented.

Another challenge is the depth of the EHR. Since the structured data starts in 2013, participants who consented in 2013 will have limited data predating their consent.

Despite these limitations, the SPS Biobank offers a unique opportunity to conduct research that has real-world impacts and can improve a multitude of health outcomes for Latinos. By continuing to follow up with participants over time, the SPS Biobank can contribute to a more comprehensive understanding of the health challenges faced by Latinos and help inform interventions to address these challenges clinically.

CONCLUSIONS

The SPS Biobank is a valuable resource for clinical research aimed at improving the health of the Latino community. By leveraging the expertise of academic institutions (ie, Mayo Clinic, ASU) and community care providers from Mountain Park Health Center, the biobank represents a novel approach to precision medicine initiatives that incorporates genomic data analysis and the return of results to participants and their healthcare providers.

These data and samples collected through the SPS Biobank are available for use in a wide range of research studies, with ongoing efforts to collect clinical and self-reported medical information that will continue to enhance the value of this resource over time. As we progress and build on the strengths of the biobank, we look forward to new discoveries and insights that will improve the health and well-being of Latinos and, as a result, other underserved communities.

COLLABORATION

All SPS data collected are stored internally at Mayo Clinic and handled confidentially. Access to the SPS Biobank can be requested by contacting at spsbiobank@mayo.edu. The access to SPS Biobank for a non-Mayo Clinic investigator can only be granted if such investigator partners with a Mayo Clinic researcher.

If you cannot identify one Mayo Clinic-based investigator, we can find one for you. Please contact us at spsbiobank@mayo.edu. For additional information, please visit https://www.mayoclinic.org/sangre-por-salud-biobank or https://www.mayoclinic.org/es/sangre-por-salud-biobank for information in Spanish. Researchers in biomedical science, particularly those focused on Latino populations, are encouraged to inquire about potential access to the biorepository's resources. Future research comparing Latino cohorts within and outside the USA is also welcomed.

Author affiliations
[1]Center for Individualized Medicine, Mayo Clinic Arizona, Scottsdale, Arizona, USA
[2]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, New York, USA
[3]Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA
[4]Mountain Park Health Center, Phoenix, Arizona, USA
[5]Biomedical Ethics Research Program, Mayo Clinic, Rochester, Minnesota, USA
[6]College of Health Solutions, Arizona State University, Phoenix, Arizona, USA
[7]Division of Endocrinology, Diabetes and Metabolism, University of Arizona Medical Center, Tucson, Arizona, USA
[8]Division of Laboratory Genetics, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA
[9]Division of Endocrinology, Diabetes and Metabolism, Mayo Clinic Arizona, Scottsdale, Arizona, USA

Acknowledgements The authors wish to acknowledge the institutional leadership at Mayo Clinic for their long-standing support of the project. We also thank the study coordinators for their efforts in recruiting the participants, and the programme coordinators for their daily efforts in coordinating and supervising the study. We especially want to thank the internal medicine providers at Mountain Park Health Center for their collaborative spirit and willingness to have their patients participate in this effort. We acknowledge and thank the Community Advisory Board for their contributions to this work. Lastly, we thank all participants for their willingness to complete this project.

Contributors EDF, DS, GQS, LM, ST and JRC jointly conceived and coordinated the research project. EDF and EZ wrote the manuscript. ER and MEG in collaboration with CG handled the data management/transfer from Mountain Park Health Center to Mayo Clinic. Statistical analysis was completed mostly by MEG with valuable input from ER, while CG KM and RRS helped with the involvement with the Community Advisory Board. GQS, EZ, CG, VH, GMG and DBP coordinated the Community Advisory Board activity. GMG, VH, SL, DBP, EZ and CG supervised the recruitment of all participants. SB contributed to the GWAS data analysis. MEG, ER, SB, VH, KM, RRS, GQS and JRC critically revised the manuscript. EDF/EDF is responsible for the overall content as guarantor. All authors approved the final draft.

Funding Sangre por Salud Biobank was funded by the Center for Individualized Medicine.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Mayo Clinic Institutional review Board, ID 12-008503. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Access to the SPS Biobank can be requested by contacting at spsbiobank@mayo.edu. Data requests will be evaluated on an individual basis and undergo eh review process for approval requests as defined by the SPS Core Committee.

ORCID iDs
Suzette Bielinski http://orcid.org/0000-0002-2905-5430
Elena De Filippis http://orcid.org/0000-0002-5622-6709

REFERENCES

1 Krogstad JM, Lopez HM. Hispanic nativity shift. 2014.
2 Bureau USC. United states census data. 2021.
3 Turner BE, Steinberg JR, Weeks BT, *et al*. Race/ethnicity reporting and representation in US clinical trials: a cohort study. *Lancet Reg Health Am* 2022;11:100252.
4 Hussain-Gambles M, Atkin K, Leese B. Why ethnic minority groups are under-represented in clinical trials: a review of the literature. *Health Soc Care Community* 2004;12:382–8.
5 Rochon PA, Mashari A, Cohen A, *et al*. The inclusion of minority groups in clinical trials: problems of under representation and under reporting of data. *Account Res* 2004;11:215–23.
6 Avilés-Santa ML, Hsu L, Lam TK, *et al*. Funding of Hispanic/Latino Health-Related Research by the National Institutes of Health: An Analysis of the Portfolio of Research Program Grants on Six Health Topic Areas. *Front Public Health* 2020;8:330.
7 Calderón JL, Baker RS, Fabrega H, *et al*. An ethno-medical perspective on research participation: a qualitative pilot study. *MedGenMed* 2006;8:23.
8 Ceballos RM, Knerr S, Scott MA, *et al*. Latino beliefs about biomedical research participation: a qualitative study on the U.S.-Mexico border. *J Empir Res Hum Res Ethics* 2014;9:10–21.
9 Cornejo-Torres Y, Boniface E, Lopez E, *et al*. Willingness to Participate in Health Research: Mexican and Mexican American Women's Perspectives. *Womens Health Rep (New Rochelle)* 2022;3:785–94.
10 Braveman PA, Kumanyika S, Fielding J, *et al*. Health disparities and health equity: the issue is justice. *Am J Public Health* 2011;101 Suppl 1:S149–55.
11 Arevalo M, Heredia NI, Krasny S, *et al*. Mexican-American perspectives on participation in clinical trials: A qualitative study. *Contemp Clin Trials Commun* 2016;4:52–7.
12 Ramirez AG, Chalela P. Equitable Representation of Latinos in Clinical Research Is Needed to Achieve Health Equity in Cancer Care. *JCO Oncol Pract* 2022;18:e797–804.
13 Phillips KA. *The Coming Era Of Precision Health*. . Health Affairs, 2021:40. 363.
14 Conomos MP, Laurie CA, Stilp AM, *et al*. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet* 2016;98:165–84.
15 Moreno-Estrada A, Gignoux CR, Fernández-López JC, *et al*. Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 2014;344:1280–5.
16 Manichaikul A, Palmas W, Rodriguez CJ, *et al*. Population structure of Hispanics in the United States: the multi-ethnic study of atherosclerosis. *PLoS Genet* 2012;8:e1002640.
17 National Academies of Sciences E, Medicine, Division of B. The national academies collection: reports funded by national institutes of health. In: *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. Washington (DC): National Academies Press (US), 2023.
18 Shaibi G, Singh D, De Filippis E, *et al*. The Sangre Por Salud Biobank: Facilitating Genetic Research in an Underrepresented Latino Community. *Public Health Genomics* 2016;19:229–38.
19 Olson JE, Ryu E, Hathcock MA, *et al*. Characteristics and utilisation of the Mayo Clinic Biobank, a clinic-based prospective collection in the USA: cohort profile. *BMJ Open* 2019;9:e032707.
20 American Diabetes Association. Standards of Medical Care in Diabetes—2013. *Diabetes Care* 2013;36:S11–66.

21 Goodman RA, Posner SF, Huang ES, *et al*. Defining and measuring chronic conditions: imperatives for research, policy, program, and practice. *Prev Chronic Dis* 2013;10:E66.
22 Standage-Beier CS, Garcia LA, De Filippis E, *et al*. Association of Vitamin D Receptor Gene Polymorphisms with Cardiometabolic Phenotypes in Hispanics: A Life Course Approach. *Nutrients* 2023;15:2118.
23 Arora A, Jack K, Kumar AV, *et al*. Genome-Wide Association Study of Gallstone Disease Identifies Novel Candidate Genomic Variants in a Latino Community of Southwest USA. *J Racial Ethn Health Disparities* 2025;12:234–40.
24 Everhart JE, Khare M, Hill M, *et al*. Prevalence and ethnic differences in gallbladder disease in the United States. *Gastroenterology* 1999;117:632–9.
25 Ferkingstad E, Oddsson A, Gretarsdottir S, *et al*. Genome-wide association meta-analysis yields 20 loci associated with gallstone disease. *Nat Commun* 2018;9:5101.
26 Puppala S, Dodd GD, Fowler S, *et al*. A genomewide search finds major susceptibility loci for gallbladder disease on chromosome 1 in Mexican Americans. *Am J Hum Genet* 2006;78:377–92.
27 Shaibi GQ, Kullo IJ, Singh DP, *et al*. Developing a Process for Returning Medically Actionable Genomic Variants to Latino Patients in a Federally Qualified Health Center. *Public Health Genomics* 2019;21:77–84.
28 Shaibi GQ, Kullo IJ, Singh DP, *et al*. Returning genomic results in a Federally Qualified Health Center: the intersection of precision medicine and social determinants of health. *Genet Med* 2020;22:1552–9.
29 Schneiderman N, Llabre M, Cowie CC, *et al*. Prevalence of diabetes among Hispanics/Latinos from diverse backgrounds: the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Diabetes Care* 2014;37:2233–9.
30 Dugani SB, Girardo ME, De Filippis E, *et al*. Risk Factors and Wellness Measures Associated with Prediabetes and Newly Diagnosed Type 2 Diabetes Mellitus in Hispanic Adults. *Metab Syndr Relat Disord* 2021;19:180–9.