nature portfolio

Corresponding author(s): Stanley L. Hazen

Last updated by author(s): 01/01/2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Untargeted metabolomics data was collected using ChromaTOF 2.32 (Leco) software. LC/MS/MS analysis was performed by LabSolution 5.91 software (Shimadzu). Flow cytometry data was acquired using FACSDiva Software (v.9.0) (BD Biosciences). Data for the in vivo thrombosis were collected by Streampix 7 - Multiple Camera DVR Software (NorPix Inc) (commercial software).
Data analysis	All data were organized and analyzed by Microsoft Excel v2013, R 4.0.2. (Vienna, Austria) (open source) with packages rms (6.2-0), survival (3.2-13), ggplot2 (3.3.3), tidyr (1.1.3), GraphPad Prism V. 9.0 (commercial software), FACSDiva Software (v.9.0) (BD Biosciences) and PASS 2022 software (Power Analysis & Sample Size), Image Pro plus software v7.0.0 (Media Cybernetics, Rockville, Maryland, USA). Custom R codes used in this manuscript are available at https://doi.org/10.5281/zenodo.6780497

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

There are restrictions to the availability of some of the clinical data generated in the present study (Fig. 1 and 2), because we do not have permission in our informed consent from research subjects to share data outside our institution without their authorizations. Where permissible, the datasets generated and/or analyzed during the present studies are available from the corresponding author (Stanley L. Hazen ,hazens@ccf.org) on request. A response can be expected within 14 days.

Custom R codes used in this manuscript are available at https://doi.org/10.5281/zenodo.6780497

The BinBase data base is accessible using the following link https://bitbucket.org/fiehnlab/binbase/src/master/?

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	In our clinical cohorts, gender is reported based on subject self-reporting. Numbers of male and female participants are shown in the main manuscript and suppl. material. (US cohort 773 females and 1376 males, in the European Cohort 249 females and 584 males). Healthy volunteers that donated blood samples for in vitro platelet assays included both males (30) and females (25). At the time of study design we did not expect gender-specific effects for our clinical observational studies. In our sensitivity analyses of the observational studies, we stratified by gender and observed a significant association of plasma erythritol levels with MACE in both males and females alike with no effect of gender on the reported association between erythritol and MACE. (Fig. 2). Gender-stratified Cox models for event risk are provided in the Supplementary Material (Table S7, S8, S9) and showed similar results for both genders with point estimates for the HR for MACE to the right of unity. We therefore did not plan to investigate gender-specific effects in our intervention study a priori when the study was designed. However, we included both male and female participants (3 males and 5 females). All subjects showed a significant increase in erythritol plasma levels following ingestion.
Population characteristics	Patient characteristics for all our 3 prospective cohort studies are detailed in table 1 and S1, S2, S3.
Recruitment	Participants for GeneBank were recruited at time of visit to the Cleveland Clinic for elective diagnostic cardiac evaluations and risk assessments. Participant for the European Cohort (LipidCardio) were recruited at the time of visit to Charité University hospital undergoing elective diagnostic coronary angiography. Clinical, gender or medical phenotypes were not considers at time of enrollment or in sample selection. Since patients in our observational cohorts were recruited at quaternary referral centers and show a high prevalence of CVD and traditional risk factors, the translationability of our findings to the general population needs to be determined. Patients living nearby the Cleveland Clinic and Charité University hospital may have been more likely to paticipate (i.e. self-selection bias). However, since the Cleveland Clinic and Charite University hospital are quaternary referral centers with a large catchment area, many patients are being referred from centers throughout the US or Europe, respectively, so that a potential self-selection bias was reduced. For the Cosette intervention study, subjects were recruited from the Cleveland, OH area. Subjects were recruited by several methods, including (but not limited to) posted advertisements, chart review, and personal interactions with patients in the cardiology outpatient departments at the Cleveland Clinic Foundation. All advertising materials used for this study were approved by the Cleveland Clinic IRB.
Ethics oversight	For all studies, the clinical study protocols and informed consent for human subjects were approved by the Cleveland Clinic Institutional Review Board or the local research ethics committee at the Charité-Universitätsmedizin (GeneBank IRB 4265, European Validation Cohort EA1/135/16, COSETTE IRB 21-005, healthy volunteer blood donors for platelet related studies IRB 09-506). Written informed consent was obtained from all subjects.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All experiments were performed at least in triplicate or where otherwise noted, with a larger sample size. Results shown often reflect data summed from multiple experiments with cumulatively larger sample sizes. Power calculations for in vivo thrombosis studies suggested a group size of at least 8 animals (based on an expected SD (of aggregation responses) of 18% for the assay to detect a change of 30% or more in the means, to a power of 0.9 with a type 1 error rate of 5%). We used the following online tool for the power analysis: http://powerandsamplesize.com/Calculators/Compare-2-Means/2-Sample-Equality. For the COSETTE Intervention study, power calculations suggested a samples size of 10 per group with an expected SD (of erythritol levels) of approximately 20% for postprandial erythritol levels (based on previous literature) to detect a change in 30% or more in the means, to a power of 0.9 with a type 1 error rate of 5%). We used the following online tool for the power analysis: http://powerandsamplesize.com/Calculators/Compare-2-Means/2-Sample-Equality. The first 2 subjects enrolled in the erythritol pharmacokinetics studies were followed up to 12 days after the erythritol intervention as per study protocol. After we noticed that erythritol levels were back to baseline after 7d, we streamlined our protocol in the interest of being both practical and less demanding on subjects, and sampled blood only for out to 7d following ingestion. Since the first 8 subjects all showed similar post prandial erythritol peak levels and time course of elimination, we did not enroll more volunteers. For the in vitro platelet assays no power calculation was performed. Numbers were based on past experiences with platelet-related studies. The sample sizes for the Discovery Cohort and US Validation Cohort were based on previous observations with biomarkers for cardiovascular risk prediction in GeneBank and no power calculation was performed a priori. A post-hoc sample size determination performed with PASS software 2022 (Power Analysis & Sample Size) for the US Validation Cohort based on the variation of erythritol levels seen in the Discovery Cohort estimated that a minimal overall sample size of 656 subjects (of which 162 are in the first quartile and 162 are in the fourth quartile) is needed to detect a difference between the first quartile and the fourth quartile with a power of 80% at a 0.05 significance level assuming a hazard ratio of 2.5 (Q4 vs Q1). For these calculations, we also anticipated that the proportions of subjects having a MACE during the study is 0.06 for the first quartile group and 0.2 for the fourth quartile group. These results assume that the hazard ratio is constant throughout the study and that Cox proportional hazards regression is used to analyze the data. For the US Validation Cohort a subset (n=2,149) sequential consenting participants from GeneBank (clinicaltrials.gov number: NCT00590200) was used. Since the measurement of erythritol is labor-intensive we did not analyze erythritol levels in all study participants of GeneBank (n >10,000) and sample size was based on previous observations with biomarkers for cardiovascular risk prediction in GeneBank (see above paragraph). All investigators performing laboratory analyses of patient samples were blinded to patient clinical and demographic data. For the European Validation Cohort, only 833 were available and in all samples erythritol was analysed using LC/MS/MS without exclusion. Therefore, no sample size calculation was performed. All investigators performing laboratory analyses of patient samples were blinded to patient clinical and demographic data.
Data exclusions	No human sample data was excluded from the analyses presented. There were no mice excluded.
Replication	Clinical associations reported in our untargeted metabolomics studies (Discovery Cohort, n=1,157) were extensively replicated using targeted metabolomics studies in 2 independent large clinical cohorts: Stable isotope dilution liquid chromatography tandem mass spectrometry (LC/MS/MS) was used to quantify erythritol in serum samples from a non-overlapping cohort of independent subjects (US Cohort, n=2,149) from GeneBank at the Cleveland Clinic (clinicaltrials.gov number: NCT0059020). In a third clinical study (the European Cohort, n=833), serum erythritol levels were quantified by stable isotope dilution LC/MS/MS in samples from sequential patients undergoing elective diagnostic coronary angiography enrolled in the observational LipidCardio study at the quaternary referral center, Charité University hospital, Campus Benjamin Franklin. Replicate experiments were performed for all studies as noted in figure legends, methods, and Source Data. Experimental findings were reliably reproduced
Randomization	Mice in all studies were randomized to their particular groups at time of allocation to experimental groups. Three observational clinical studies were performed. The first (discovery) and second (non-overlapping validation) analyses were performed on sequential stable subjects without evidence of acute coronary syndrome who underwent elective diagnostic coronary angiography (cardiac catheterization or coronary computed tomography) for evaluation of coronary artery disease (CAD) within the GeneBank Cohort, which was enrolled at the quaternary referral center, Cleveland Clinic, Cleveland, Ohio, USA (clinicaltrials.gov number: NCT00590200). A third clinical study involves a distinct validation cohort. Here, samples analyzed were from sequential patients undergoing elective diagnostic coronary angiography enrolled in the observational LipidCardio study at the quaternary referral center, Charité University hospital, Campus Benjamin Franklin, Berlin, Germany (German Clinical Trial Register (drks.de); Identifier: DRKS00020915). The interventional study Cosette (clinicaltrials.gov number: NCT04731363) is a non-randomized study wherein each consented volunteer undergoes a series of studies at baseline (prior to intervention) and following intervention at prescribed times. Since we compared baseline levels of erythritol with postprandial levels each subject was their own control and a non-randomized design performed. For experiments other than mice studies, participants/samples for in vitro/ ex vivo studies were randomly allocated to experimental groups/ treatments.
Blinding	All investigators performing laboratory analyses of patient samples were blinded to patient clinical and demographic data. Investigators performing quantitative analyses of thrombus formation, and metabolites levels in mouse plasma were blinded to group allocation with samples labeled by code only. Laboratory personnel performing all laboratory analyses in samples related to the interventional study (COSETTE) were blinded to sample group allocation and clinical data during analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

All assays included appropriate positive and negative controls as indicated, for assay validation. If applicable, the assays were designed per the manufacturer's instructions.
Alexa Fluor® 488-conjugated anti-human CD42b antibody (catalogue # 303914, Biolegend, San Diego, CA, USA)
anti-P-selectin (CD62P-PE, Catalogue # 555524, BD PharMingen, San Diego, CA, USA)
FITC conjugated PAC1, Catalogue #3 40507, BD PharMingen, San Diego, CA, USA)
PE IgG isotype control, Catalogue # 555749 BD PharMingen, San Diego, CA, USA)
FITC IgM Isotype control, Catalogue # 555583, BD PharMingen, San Diego, CA, USA)

Validation

All purchased antibodies were validated by their manufacturers and further in-house testing.
BioLegend, https://www.biolegend.com/en-us/reproducibility
The reproducibility of published research has emerged as an urgent topic in today's scientific community. From funding agencies to researchers to manufacturers and publishers, it is critical for all of these groups to align themselves to ensure that research is done with rigor and is reproducible. How has BioLegend stepped up to meet these needs? In addition to sponsoring and collaborating with the Global Biological Standards Institute (GBSI) on setting antibody standards, we undertake extensive measures to ensure quality products that meet reproducibility requirements today and into the future. BioLegend spends a considerable amount of effort in creating new antibodies for research. The majority of these new antibody products are monoclonal antibodies (mAb) produced from hybridomas. Clones of these hybridomas are carefully selected based on a number of criteria including robust growth and efficient production of a single clone of antibody that is specific to the intended target. The best clones move on to applications testing. All newly developed clones at BioLegend undergo validation testing for multiple applications. This serves as a cross-check for specificity and provides clarity for research uses. Maintaining lot-to-lot consistency is vital for reproducibility. It simply is not enough for antibody manufacturers to validate antibodies just once. Pass/fail specification requirements are essential for quality control testing of every lot of product. BioLegend maintains records for all lotspecific testing

The Alexa Fluor® 488-conjugated anti-human CD42b antibody (catalogue # 303914, Biolegend, San Diego, CA, USA) was quality tested for flow cytometry and reactivity to human samples veryfied by the manufacturer: https://d1spbj2x7qk4bg.cloudfront.net/en-us/products/alexa-fluor-488-anti-human-cd42b-antibody-13189?pdf=true&displayInline=true&leftRightMargin=15&topBottomMargin=15&filename=Alexa%20Fluor%C2%AE%20488%20anti-human%20CD42b%20Antibody.pdf&v=20220829083428

BD Biosciences, https://www.biocompare.com/Antibody-Manufacturing/355107-Antibody-Manufacturing-Perspectives-BDBioscience/
We conduct quality control (QC) testing in primary model systems to ensure biological accuracy in an ISO 9001 certified facility. BD carefully selects and characterizes antibody content in product development and tests in relevant primary model systems to ensure biological accuracy. BD conducts rigorous QC testing of each antibody lot tested side-by-side with a previously produced lot as reference. Our product development process includes testing on a combination of primary cells, cell lines and/or transfectant cell models with relevant controls using multiple immunoassays to ensure biological accuracy. We also perform multiplexing with additional antibodies to interrogate antibody staining in multiple cell populations. BD believes antibody validation is critical to ensure accurate scientific results. Both the consumer and the reagent provider share the responsibility for reproducible science.

The anti-P-selectin (CD62P-PE, Catalogue # 555524, BD PharMingen, San Diego, CA, USA) was quality tested for flow cytometry by the manufacturer and human reactivity veryfied: https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/pe-mouse-anti-human-cd62p.555524

The FITC-conjugated PAC1 antibody (Catalogue #340507, BD PharMingen, San Diego, CA, USA) was quality tested for flow cytometry confirmed to have human reactivity by the manufacturer: https://www.bdbiosciences.com/en-ca/products/reagents/flow-cytometry-reagents/clinical-discovery-research/single-color-antibodies-ruo-gmp/fitc-mouse-anti-human-pac-1.340507

In addition, the P-selectin and PAC1 antibodies were validated by the investigators using both unstimulated and ADP-stimulated human platelets and the appropriate dilutions for the experiments determined by titration.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Male and female C57BL/6J mice, 12-14 weeks old, were purchased from The Jackson Laboratory and maintained in our facilities for in vivo experiments. Animals were maintained under a strict 14 h light/10 h dark cycle with water and food available ad libitum and temperature of 20-26 °C and 30-70% humidity.
Wild animals	No wild animals were used.
Reporting on sex	Based on our results from the observational studies outlined in the manuscript (Fig. 2 and Supplemantary Material) we did not expect gender-specific effects on erythritol-dependent enhancement in in vivo thrombosis a priori when we designed these studies. However, we included both male and female mice into the studies. 15 female and 16 male mice were included in the analyses reported. We observed erythritol-dependent enhancement in the in vivo thrombosis formation in both male and female mice.
Field-collected samples	No field-collected samples were used.
Ethics oversight	All animal model studies were approved by the Institutional Animal Care and Use Committee at the Cleveland Clinic.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	All prospective cohort studies and the prospective intervention study in our manuscript were registered at publicly accessible databases: GeneBank is registered at clinicaltrials.gov, identifier: NCT00590200, the European Validation Cohort (LipidCardio) is registered under German Clinical Trial Register (drks.de); Identifier: DRKS00020915. The Cosette Intervention study is registered at clinicaltrials.gov, identifier: NCT04731363).
Study protocol	Information about the studies used in this manuscript can be accessed at clinicaltrials.gov or drks.de. The study protocol for the Erythritol intervention study (COSETTE) is provided for details. The data used in the manuscript includes erythritol measurement after ingestion of 30g of erythritol as outlined in the study protocol. For the measurement of erythritol levels after ingestion, we established a time course up to 12d for the first 2 subjects and based on the results streamlined the protocol to only include serial measurements up to 7 days after ingestion in the interest of our study subjects (upon 7d after exposure the erythritol levels were similar to baseline). Since all 8 subjects showed similar postprandial erythritol peak levels and elimination over time, we enrolled 8 subjects and the reported data represent the final analysis. The registration of COSETTE focuses on the 30 min time point after ingestion. The registration at Clinicaltrials.gov and the IRB protocol for COSETTE also includes a distinct non-overlapping second separate study that assesses platelet functional changes after ingestion. These studies are ongoing, and not part of this manuscript.
Data collection	All subjects were enrolled at either the main campus of the Cleveland Clinic (GeneBank) or at the Charité-Universitätsmedizin Berlin. The discovery and US validation cohort consisted of sequential stable adult patients undergoing cardiac risk assessment for symptom evaluation at a quaternary referral center between 2001-2007 with MACE outcomes adjudicated up to 3 years. The European validation cohort (LipidCardio study) consists of sequential patients that were enrolled between 2016-2018 at the quaternary referral center, Charité University hospital, Campus Benjamin Franklin with a follow-up for 3 years. Blood samples were collected in a private setting by a research nurse or coordinator trained in phlebotomy. Demographic data was participant-reported. Other test results (e.g. characterizations) were obtained from a review of the participants' medical records.
Outcomes	Major adverse cardiovascular events were predefined as non-fatal myocardial infarction, stroke or death. All follow-up data was collected through phone calls with participants and/or manual chart review (review of their electronic medical record), information from family members, and Social Security Death Index query.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Washed platelets were separated by centrifugation at 500 x g for 10 min and re-suspended in modified Hank's buffered salt solution without PGE1. Platelet suspensions (100mL; 2x108 platelets/mL) stimulated with 2 µM ADP for 10 min and incubated with primary antibodies in the dark for 20 min. The platelet suspensions were then fixed with 100 mL of 2% paraformaldehyde.
Instrument	Data was acquired on a flow cytometer (FACS LSR Fortessa, BD Biosciences, Franklin Lakes, New Jersey, USA).
Software	Data was analyzed using FACSDiva Software (v.9.0) (BD Biosciences)
Cell population abundance	We did not sort platelets. Washed platelet we used and platelet purity verified by CD42 expression.
Gating strategy	Platelets were selected on a FSC-A/SSC-A plot on log-scales. Aggregates were excluded on a SSC-H (log)/SSC-W (linear) plot.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.