**Materials Design Analysis Reporting (MDAR)**

**Checklist for Authors**

The [MDAR framework](https://osf.io/xfpn4/) establishes a minimum set of requirements in transparent reporting mainly applicable to studies in the life sciences.

*eLife* asks authors to **provide detailed information within their article** to facilitate the interpretation and replication of their work. Authors can also upload supporting materials to comply with relevant reporting guidelines for health-related research (see [EQUATOR Network](http://www.equator-network.org/%20)), life science research (see the [BioSharing Information Resource](http://biosharing.org/)), or animal research (see the [ARRIVE Guidelines](http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1000412) and the [STRANGE Framework](https://doi.org/10.1038/d41586-020-01751-5); for details, see *eLife*’s [Journal Policies](https://reviewer.elifesciences.org/author-guide/journal-policies)). Where applicable, authors should refer to any relevant reporting standards materials in this form.

For all that apply, please note **where in the article** the information is provided. Please note that we also collect information about data availability and ethics in the submission form.

**Materials:**

|  |  |  |
| --- | --- | --- |
| **Newly created materials** | **Indicate where provided: section/figure legend** | **N/A** |
| The manuscript includes a dedicated "materials availability statement" providing transparent disclosure about availability of newly created materials including details on how materials can be accessed and describing any restrictions on access. | The manuscript details a protocol outlining a collaborative analysis involving multiple studies sharing common components. The data is not accessible, as it has yet to be made available for analysis. The data collection is ongoing. The data will be loaned to the NCI. Access to this data will require individual requests to the respective sites. As a consortium, we aim to furnish essential information to enable verification and replication upon approval from the individual sites.  The primary objective of this study is to develop AI algorithms that will undergo validation across various datasets within the consortium. Ultimately, our aim is to make these algorithms publicly accessible once they are finalized. |  |
|  |  |  |
| **Antibodies** | **Indicate where provided: section/figure legend** | **N/A** |
| For commercial reagents, provide supplier name, catalogue number and [RRID](https://scicrunch.org/resources), if available. |  | x |
|  |  |  |
| **DNA and RNA sequences** | **Indicate where provided: section/figure legend** | **N/A** |
| Short novel DNA or RNA including primers, probes: Sequences should be included or deposited in a public repository. | We are using a PCR based technology to detect HPV DNA. We use a commercially available technology, ScrennFire from Atila Biosystems and the sequences are not available |  |
|  |  |  |
| **Cell materials** | **Indicate where provided: section/figure legend** | **N/A** |
| Cell lines: Provide species information, strain. Provide accession number in repository OR supplier name, catalog number, clone number, OR RRID. |  | x |
| Primary cultures: Provide species, strain, sex of origin, genetic modification status. |  | x |
|  |  |  |
| **Experimental animals** | **Indicate where provided: section/figure legend** | **N/A** |
| Laboratory animals or Model organisms: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository OR supplier name, catalog number, clone number, OR RRID. |  | x |
| Animal observed in or captured from the field: Provide species, sex, and age where possible. |  | x |
|  |  |  |
| **Plants and microbes** | **Indicate where provided: section/figure legend** | **N/A** |
| Plants: provide species and strain, ecotype and cultivar where relevant, unique accession number if available, and source (including location for collected wild specimens). |  | x |
| Microbes: provide species and strain, unique accession number if available, and source. |  | x |
|  |  |  |
| **Human research participants** | **Indicate where provided: section/figure legend) or state if these demographics were not collected** | **N/A** |
| If collected and within the bounds of privacy constraints report on age, sex, gender and ethnicity for all study participants. | This is a Consortium protocol. No data are yet available. The study is being performed in women in the age range of 25-49 and will include a diverse ethnicity as the study includes populations from Latin America, sub-Saharan Africa and Cambodia for Asia. |  |

**Design:**

|  |  |  |
| --- | --- | --- |
| **Study protocol** | **Indicate where provided: section/figure legend** | **N/A** |
| If the study protocol has been pre-registered, provide DOI. For clinical trials, provide the trial registration number OR cite DOI. | This is a Consortium protocol. After consultation it is not considered a clinical trial |  |
|  |  |  |
| **Laboratory protocol** | **Indicate where provided: section/figure legend** | **N/A** |
| Provide DOI OR other citation details if detailed step-by-step protocols are available. | See reference by Inturrisi et al. 2023 |  |
|  |  |  |
| **Experimental study design (statistics details) \*** | | |
| **For in vivo studies: State whether and how the following have been done** | **Indicate where provided: section/figure legend. If it could have been done, but was not, write “not done”** | **N/A** |
| Sample size determination | An initial aim is to enroll 100,000 with the expectation to have sufficient number of CIN3 cases to have enough outcomes (about 1% of the screened population). This is a conservative estimate as some of the areas may have a higher prevalence of CIN3 |  |
| Randomisation | No randomization, but whenever there is an ongoing standard of care, a comparison with the new PAVE strategy will be provided |  |
| Blinding | During the initial phase, known as the efficacy study, participants will be managed in accordance with the standard care protocols at each site. Providers will not have access to the AI score for quality or diagnosis until the validation of the strategy is confirmed. In the subsequent phase, providers will have the opportunity to compare their own approach with our strategy. Nevertheless, any alterations in clinical management will adhere to the guidelines outlined by the Ministry of Health (MoH). |  |
| Inclusion/exclusion criteria | All defined at the recruitment section |  |
|  |  |  |
| **Sample definition and in-laboratory replication** | **Indicate where provided: section/figure legend** | **N/A** |
| State number of times the experiment was replicated in the laboratory. | HPV tests are replicated in around 10% of the samples or until results are considered consistent. |  |
| Define whether data describe technical or biological replicates. | The AI functions will be evaluated in different sets, including external sets to confirm their portability. Once the algorithms are validated we aim to provide them as an open source for public use in adequate conditions. |  |
|  |  |  |
| **Ethics** | **Indicate where provided: section/submission form** | **N/A** |
| Studies involving human participants: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval. | **All sites participating in the Consortium have their IRB approval and those have been checked at NCI** |  |
| Studies involving experimental animals: State details of authority granting ethics approval (IRB or equivalent committee(s), provide reference number for approval. | IRB at the different countries have provided the individual study´s approval. |  |
| Studies involving specimen and field samples: State if relevant permits obtained, provide details of authority approving study; if none were required, explain why. | All sites require their own ethical approval, which has been attested by the NCI´s PI. Data is compiled at NCI for analysis but no specimens or samples are shipped at NCI derived from this Consortium |  |
|  |  |  |
| **Dual Use Research of Concern (DURC)** | **Indicate where provided: section/submission form** | **N/A** |
| If study is subject to dual use research of concern regulations, state the authority granting approval and reference number for the regulatory approval. |  | x |

**Analysis:**

|  |  |  |
| --- | --- | --- |
| **Attrition** | **Indicate where provided: section/figure legend** | **N/A** |
| Describe whether exclusion criteria were pre-established. Report if sample or data points were omitted from analysis. If yes, report if this was due to attrition or intentional exclusion and provide justification. | Inclusion and exclusion criteria are specified in the enrollment section |  |
|  |  |  |
| **Statistics** | **Indicate where provided: section/figure legend** | **N/A** |
| Describe statistical tests used and justify choice of tests. | See the statistical section and Egemen et al. 2023 |  |
|  |  |  |
| **Data availability** | **Indicate where provided: section/submission form** | **N/A** |
| For newly created and reused datasets, the manuscript includes a data availability statement that provides details for access (or notes restrictions on access). | Data are not available yet, and therefore, the manuscript does not contain any dataset for analysis. Once the data is generated and analyzed we will provide details for access. A site-specific for PAVE is being generated at NCI website, where requests will be processed. |  |
| When newly created datasets are publicly available, provide accession number in repository OR DOI and licensing details where available. | Once the data is generated and analyzed we will provide details for access. A site-specific for PAVE is being generated at NCI website, where requests will be processed. |  |
| If reused data is publicly available provide accession number in repository OR DOI, OR URL, OR citation. | We are working on having the URL within the NCI website for the PAVE study |  |
|  |  |  |
| **Code availability** | **Indicate where provided: section/figure legend** | **N/A** |
| For any computer code/software/mathematical algorithms essential for replicating the main findings of the study, whether newly generated or re-used, the manuscript includes a data availability statement that provides details for access or notes restrictions. | Final algorithms are not yet available but will be provided when presenting the results to allow replication. The aim is to generate algorithms that can be of public use. |  |
| Where newly generated code is publicly available, provide accession number in repository, OR DOI OR URL and licensing details where available. State any restrictions on code availability or accessibility. | We cannot yet provide an accession number as data are not available, however, we are finalizing a website within NCI where accession number in repository will be stated once the Consortium analysis has been properly evaluated and published. The Consortium will have to agree on the external sharing of full or partial data. |  |
| If reused code is publicly available provide accession number in repository OR DOI OR URL, OR citation. | We are working on having the URL within the NCI website for the PAVE study |  |

**Reporting:**

The MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives.

|  |  |  |
| --- | --- | --- |
| **Adherence to community standards** | **Indicate where provided: section/figure legend** | **N/A** |
| State if relevant guidelines (e.g., ICMJE, MIBBI, ARRIVE, STRANGE) have been followed, and whether a checklist (e.g., CONSORT, PRISMA, ARRIVE) is provided with the manuscript. | We have used the SPIRIT-AI-Checklist  It can be provided upon request |  |

\* We provide the following guidance regarding transparent reporting and statistics; we also refer authors to [Ten common statistical mistakes to watch out for when writing or reviewing a manuscript](https://doi.org/10.7554/eLife.48175).

**Sample-size estimation**

* You should state whether an appropriate sample size was computed when the study was being designed
* You should state the statistical method of sample size computation and any required assumptions
* If no explicit power analysis was used, you should describe how you decided what sample (replicate) size (number) to use

**Replicates**

* You should report how often each experiment was performed
* You should include a definition of biological versus technical replication
* The data obtained should be provided and sufficient information should be provided to indicate the number of independent biological and/or technical replicates
* If you encountered any outliers, you should describe how these were handled
* Criteria for exclusion/inclusion of data should be clearly stated
* High-throughput sequence data should be uploaded before submission, with a private link for reviewers provided (these are available from both GEO and ArrayExpress)

**Statistical reporting**

* Statistical analysis methods should be described and justified
* Raw data should be presented in figures whenever informative to do so (typically when N per group is less than 10)
* For each experiment, you should identify the statistical tests used, exact values of N, definitions of center, methods of multiple test correction, and dispersion and precision measures (e.g., mean, median, SD, SEM, confidence intervals; and, for the major substantive results, a measure of effect size (e.g., Pearson's r, Cohen's d)
* Report exact p-values wherever possible alongside the summary statistics and 95% confidence intervals. These should be reported for all key questions and not only when the p-value is less than 0.05.

**Group allocation**

* Indicate how samples were allocated into experimental groups (in the case of clinical studies, please specify allocation to treatment method); if randomization was used, please also state if restricted randomization was applied
* Indicate if masking was used during group allocation, data collection and/or data analysis