



University  
of Exeter

# Open Research Awards 2024

## **HLS disciplines prize**

**Bethany Voller**

### **Title of case study**

**Advancing the field of multimorbidity research via openness and reproducibility practices**

#### Introduction to the case study

Multimorbidity, the co-existence of two or more long-term health conditions (LTCs), will affect two thirds of adults aged 65 and over within the next decade. The MRC-funded GEMINI (Genetic Evaluation of Multimorbidity towards INdividualisation of Interventions) collaborative aims to elucidate possible links in multimorbidity using both genetic and epidemiological approaches, advancing the field of research, and improving patient outcomes. Striving for high standards of openness and reproducibility will help us to achieve this goal.

#### Detail the practices you have used to support openness/reproducibility

We have recently submitted our first publication, for which I am one of the joint first authors; the preprint is available on medRxiv.

With input from public collaborators and workshops with people with lived experience, we identified 72 LTCs. We curated diagnostic code lists with input from multiple clinician experts to ensure accuracy and validity. Maintaining consistency in definitions across multiple healthcare systems (GP and hospital), across two countries (UK and Spain), and different code vocabularies, was a challenge. Our code lists are publicly available on GitHub.

As a group, we have developed systems to track and document our work, enabling effective collaboration. We share documented and commented code, using version controlling with GitHub where possible. We have released an R package (partialLDSC); I am involved in maintaining this package.

I have created content for, and assisted in creating a Shiny app, complementing our research with results in the form of tables and interactive figures (e.g. heatmaps and scatter plots). The app aims to present our findings in a different way, enabling researchers, patients, clinicians, and laypeople to easily explore them in more depth.

I have shared methods and discussed best practice via two multimorbidity community of practice groups.

#### Outline any barriers or challenges you encountered, and how they were handled

Condition coding was a complex and time-consuming task. We spent time rigorously defining our conditions to ensure accuracy; by making our definitions public, we hope to benefit future

research, improving efficiency and consistency in the field.

To maximise the strength of my analyses, I aimed to use as much genetic data as possible. However, genetic data are not always publicly available. Authors of relevant studies were contacted – a few agreed to share their data, but most did not reply. My findings may have ended up being stronger had more researchers made their data open access.

#### What benefits have you identified from your case?

Being open enables more efficient research; if code or data needed already exists, then it need not be repeated. This is particularly important given the complexity of studying multiple LTCs. Open research resources such as the GWAS Catalog and GitHub have been invaluable to my work in this sense, giving more time for novel research.

Being open in our research has allowed us to collaborate with others working in the same field outside of our group; we are involved in two communities of practice, with regular meetings and workshops, which have been mutually beneficial.

#### How do you plan to continue this case moving forward?

Code accompanying preprints and publications will be made public with clear versions on GitHub following publication acceptance, along with R packages for novel methods or pipelines. I will endeavour to ensure that all my code is as clear and reproducible as possible – by adding comments where helpful and ensuring comprehensible documentation.

Using our defined clinical code lists, I have performed genome-wide association studies (GWAS) in UK Biobank. These have been formatted for submission to the GWAS Catalog – a searchable, consistent, and freely available database compiling GWAS data from various sources, and will be submitted shortly to be made public.

#### Conclusions

Open research is essential in maximising the impact of scientific research, enabling wider collaboration and dissemination. I have benefitted from others following open practices, and have implemented myself, or been involved in implementing a number of open practices, that I hope will be beneficial to others.

#### Links to any materials/resources relevant to the case

medRxiv preprint: <https://doi.org/10.1101/2024.05.13.24307009>.

GEMINI GitHub: <https://github.com/GEMINI-multimorbidity/>.

partialLDSC R package, developed by a colleague, maintained by me and others:  
<https://github.com/GEMINI-multimorbidity/partialLDSC>.

GEMINI shiny app: <https://gemini-multimorbidity.shinyapps.io/atlas/>.

GWAS Catalog: <https://www.ebi.ac.uk/gwas/>.