A statistical likelihood-ratio methodology for integrating enriched case data from national laboratory testing within large-scale case-control analyses (PS4)

Sophie Allen, Charlie F. Rowlands, Alice Garrett, Fergus Couch, Marcy Richardson, Tina Pesaran, Jo Pethick, Katrina Lavelle, Fiona E. McRonald, Sally Vernon, Miranda Durkie, George J. Burghel, Alison Callaway, Rachel Robinson, Joanne Field, Bethan Frugtniet, Sheila Palmer-Smith, Jonathan Grant, Judith Pagan, Trudi McDevitt, Katie Snape, Helen Hanson, Terri McVeigh, Steven Hardy, Clare Turnbull.

Abstract

Background: For v3.0 of the ACMG/AMP guidelines, case-control data is applied under PS4 at a 'strong' level of evidence. However, while quantitative methodologies have been developed for other evidence criteria to empower their use outside of the standard categorical evidence strengths, no such methodology has been established for case-control data. With the release of large-scale unselected case-control datasets for breast cancer (BRIDGES, CARRIERS, and UK Biobank) and expansion of nationally-collected laboratory datasets enriched for pathogenic variant carriers (data collated and held by the National Disease Registration Service, NDRS, and data collected by Ambry Genetics Corp), there is now potential to combine datasets across data collection contexts in a more quantitative manner using novel likelihood ratio tools.

Methods: Using the published PS4-LRCalc[1] tool (a flexible methodology for quantifying case-control evidence as a likelihood ratio using prespecified odds ratio thresholds), rules were iterated for combination of unselected and enriched laboratory datasets. These rules included development of 'Enrichment Factors' representing the degree by which clinically-ascertained laboratory series are enriched for pathogenic variants compared to unselected breast cancer series. Enriched laboratory datasets were paired with equivalent population controls.

Results: In total, data was combined for 11,462 missense variants from 342,760 female breast cancer patients and 671,290 controls of Western European ancestry for five breast cancer susceptibility genes (*BRCA1, BRCA2, PALB2, ATM, CHEK2).* A combined log likelihood ratio (LLR) was produced for 3,073 missense variants. Under an assumption of standard penetrance for each gene (OR≥4 for *BRCA1* and *BRCA2*, OR≥3 for *PALB2*, and OR≥2 for *ATM* and *CHEK2*), 811 variants received evidence towards pathogenicity (LLR≥1), and 1,867 received evidence towards benignity (LLR≤1).

Conclusion: This flexible, variant-level methodology incorporates nationally-collected 'enriched' datasets with unselected case-control cohorts, ultimately expanding the available information for case-control analysis, boosting power for existing analyses and enabling future exploration of reduced penetrance odds ratio thresholds.

1 Rowlands CF, Garrett A, Allen S et al. The PS4-likelihood ratio calculator: flexible allocation of evidence weighting for case-control data in variant classification. Journal of medical genetics 2024.