Statistica Sinica **33** (2023), 1-23
doi:https://doi.org/10.5705/ss.202021.0164

MENDELIAN RANDOMIZATION TEST OF CAUSAL EFFECT USING HIGH-DIMENSIONAL SUMMARY DATA

Lu Deng, William Wheeler and Kai Yu

Nankai University, Information Management Services, Silver Spring and National Cancer Institute

Abstract: Mendelian randomization (MR) uses genetic variants as instrumental variables (IVs) to assess the causal effect of a risk factor on an outcome in the presence of unmeasured confounding. There is growing interest in conducting MR analyses using summary statistics on each IV's association with the risk factor and the outcome, which are generated from large-scale genome-wide association studies (GWAS). Most existing approaches use summary data on a set of IVs that have been established as being associated with the risk factor. They often have limited power because the set of identified IVs jointly explain only a small proportion of the variation in the measure of the risk factor. We propose a new MR testing procedure that takes full advantage of summary data on tens of thousands of genetic variants studied by GWAS. The test statistic is the maximum of a sequence of modified K-statistics defined by a range of thresholds. Compared with existing approaches, this new test gains power by collecting signals from many undetected IVs throughout the genome, and is robust to both balanced and unbalanced pleiotropy. We investigate the theoretic properties of the proposed procedure and demonstrate its advantages over existing ones using simulation studies and a real example.

Key words and phrases: Genome-wide association studies, instrumental variables, mendelian randomization, pleiotropic effect, summary statistics.

1. Introduction

Mendelian randomization (MR) analysis uses genetic variants as instrumental variables (IVs) to estimate the causal effect of a risk factor on an outcome based on observational studies (Lawlor et al. (2008); Smith and Ebrahim (2003)). MR is becoming an effective tool for studying the causal relationship between a risk exposure and a disease outcome, because many robust findings on the genetic basis underlying various common traits have been accumulated through large-scale genome-wide association studies (GWAS) over the past decade.

Corresponding author: Kai Yu, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA. E-mail: yuka@mail.nih.gov.