

### **Disclosure statement by Heidi Williams**

- The two data sets used in this paper cannot be posted online. The first data set is collected by a private company (IMV; <http://www.imvinfo.com>) and is therefore proprietary, but it is available for purchase.
- The second source of data is the California Cancer Registry, which records information about cancer patients diagnosed in the state of California (<http://www.ccrca.org/>). The collection of the cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement U58DP003862-01 awarded to the California Department of Public Health.
- We obtained a research version of this database with confidential variables as part of a research agreement with PI Professor Joseph Doyle at MIT. The research agreement was approved by the California Committee for the Protection of Human Subjects and the MIT COUHES (Committee on the Use of Humans as Experimental Subjects). The state of California has a standardized application process for obtaining these data (described here: [http://www.ccrca.org/Data\\_and\\_Statistics/Cancer\\_Data\\_for\\_Research.shtml](http://www.ccrca.org/Data_and_Statistics/Cancer_Data_for_Research.shtml)).
- Financial support for this work was provided by the National Institute on Aging (Einav and Finkelstein, R01-AG032449) and National Science Foundation (Williams, 1151497).