

DEMOGRAPHIC DATA FOR LOCAL AREAS

HARLEY B. MESSINGER
INSTITUTE FOR HEALTH RESEARCH, BERKELEY

1. Data sources

1.1. *Census.* The 1970 Census offers much better access to local data than previous decennial censuses. Examples are "block-face" coding and the preparation of census tract directories. The latter permit coding street addresses to census tracts and were never before available for so many cities at one time, the former would permit obtaining data from housing along designated streets with several levels of auto traffic density for a study of air pollution effects on people living at the several levels.

1.2. *Local records.* Vital statistics records have long been popular for epidemiologic studies but recently the value of other routinely collected local items has become appreciated. Santa Clara County in the San Francisco Bay Area has linked files for different departments by coding census tract onto many county records.

1.3. *Surveys.* While much can be done with data gathered for other purposes, an epidemiologic study will usually have to collect information directly as well. Having an interviewer survey each individual household no longer is considered necessary to get valid data. Hochstim's Cervical Cytology Study in Alameda County showed how to increase efficiency of data collection by use of the telephone or the mails [2].

2. Data handling

2.1. *Data banking.* Larger studies may require huge data files meriting the label of "data bank." At one extreme, too little attention is paid to planning for this aspect and the study founders in a mass of doubtful data; at the other, the gathering of information becomes an end in itself and more is collected than can possibly be analyzed with funds available. With the growing realization that protection of confidentiality requires positive measures, even more care will be required to keep data processing budgets within bounds. In the "link file" system used by the American Council on Education for a longitudinal study [1] the identity of a questionnaire respondent is kept secret by having the file linking him to his particular set of answers stored in a foreign country.

2.2. *Record linkage.* Besides the aggregation of statistics by area, the linkage