

Epidemiologic research with administrative databases: red herrings, false alarms and pseudo-epidemics

David A. Grimes*

Department of Obstetrics and Gynecology, University of North Carolina School of Medicine, Chapel Hill, NC, USA

*Correspondence address. Tel: +1-919-656-7227; E-mail: david_grimes@med.unc.edu

Submitted on May 13, 2015; resubmitted on May 13, 2015; accepted on June 1, 2015

In 1994, Altman (1994) startled the medical world by declaring that poor research is as unethical as fraudulent research. His reasoning was simple: both kinds of research mislead which, in turn, harms patients. A decade later, others (von Elm and Egger, 2004) found that the problem had persisted. With the growing indiscriminate use of administrative databases for epidemiologic research (Grimes, 2010) the problem has only gotten worse since then. The record-linkage study of assisted reproductive technologies (ART) and cancer in this issue (Reigstad *et al.*, 2015a) illustrates the problem.

Failure to follow reporting guidelines

Research reports should follow the internationally accepted reporting guidelines (Simeria *et al.*, 2010). The STROBE (STrengthening the Reporting of OBServational Studies in Epidemiology) guidelines were developed to promote more complete and transparent reporting of observational studies (von Elm *et al.*, 2007). The report in this issue (Reigstad *et al.*, 2015a) neither mentions nor cites the STROBE guidelines. In addition, Figure 2 of the report (Reigstad *et al.*, 2015a) improperly plots hazard ratios and confidence intervals on a linear (arithmetic) scale (Levine *et al.*, 2010).

Fuzzy exposure

In any cohort study, the exposure must be clear, specific and measurable (Grimes and Schulz, 2002). In the report by Reigstad and colleagues (Reigstad *et al.*, 2015a), the exposure was a disparate group of ill-described techniques to promote fertility: the administrative database had scant information about them. The regimens used, drugs administered, number of treatment cycles, and combinations of treatments were unknown or unclear. Aggregating these heterogeneous exposures as 'ART' is analogous to combining all methods of fertility control as 'contraception' then studying cancer as an outcome. Any result is incapable of interpretation.

Inadequate control for potential confounding factors

In these Norwegian databases (Reigstad *et al.*, 2015a) information about potential confounding factors was inadequate. The databases were not designed for epidemiologic research. The only factors for which adjustments could be made were 'age at start of follow-up, parity, region of residence and calendar year at follow-up' (Reigstad *et al.*, 2015a). Attempting to study the epidemiology of specific cancers without considering known risk factors is untenable.

Many risk factors for cancer are known. Cigarette smoking (Winkelstein, 1990) and sexual behavior (Wakeham and Kavanagh, 2014) are risk factors for cervical cancer. The report from Reigstad *et al.* (2015a) had no information about smoking, number of sexual partners or use of barrier contraceptives (Lam *et al.*, 2014). Obesity is strongly related to endometrial cancer (Nevadunsky *et al.*, 2014); Reigstad *et al.* (2015a) reported no information on weight, body mass index (BMI) or other measures of adiposity. Oral contraceptive use is associated with a lower risk of ovarian cancer (Ness *et al.*, 2000) but Reigstad *et al.* (2015a) had no information on this common exposure. Trying to study the epidemiology of cancers with these Norwegian databases is analogous to trying to study the epidemiology of automobile accidents with a Department of Motor Vehicles database. Age, height, weight, and eye and hair color cannot explain why accidents occur (Grimes, 2010).

Quick-and-dirty research

Research using existing administrative databases has obvious appeal. Instead of doing a proper (and thus expensive) study of a specific *a priori* hypothesis with a written protocol, epidemiologists pore over huge data sets in search of something statistically significant, a process derisively termed 'risk-factorology.' (Smith, 2001). The more comparisons are made, the greater is the likelihood that something will pop up as statistically significant (even when no association exists). Indeed, after adjustment for the scores and scores of comparisons

made (Reigstad *et al.*, 2015a) statistical significance vanished. No safety signals remained (Silver, 2012).

Journals are enamored of large administrative databases as well. Huge numbers of patients produce narrow confidence intervals and tiny *P* values. However, because these databases were not designed for epidemiologic research, the results are often wrong: garbage in, garbage out (Grimes, 2010). Large numbers increase the precision of results but do not address the more important issue of validity. A big database study can give an invalid answer with a narrow confidence interval. Stated alternatively, the result can be precisely wrong (Lidegaard *et al.* 2012a,b). That is dangerous.

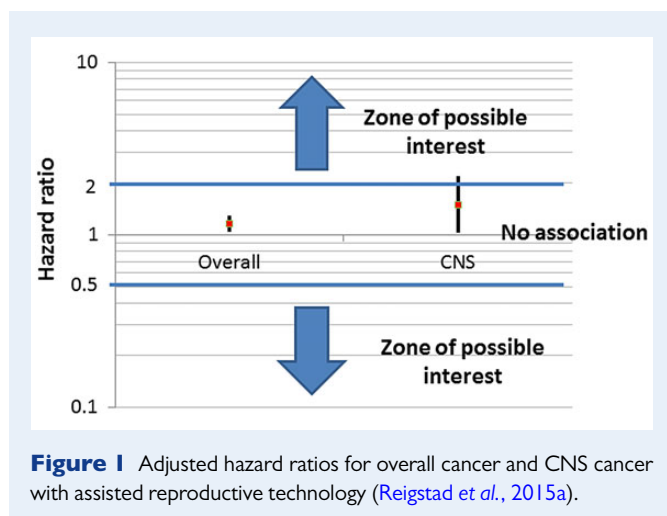
Weak associations

The need for circumspection is critical with weak associations. In Hill's classic paper on judging causal associations (Hill, 1965), the first consideration was strength of the association. In a cohort study, strength is measured by the relative risk. Hill gave two examples: cigarette smoking and death from lung cancer and polluted water and cholera. The smallest relative risks in these two examples were 8 and 14, respectively (Hill, 1965). Relative risks of that size are likely due to causation, not bias.

The lower limit of discrimination for cohort studies is widely considered to be a relative risk of about 2, or its reciprocal, 0.5 (Fig. 1) (Grimes and Schulz, 2012). Relative risks between 0.5 and 2 are likely due to bias, which is present in all observational studies. Values above 2 or below 0.5 (Fig. 1) may merit attention (Grimes and Schulz, 2012). In the report from Reigstad *et al.* (2015a) the adjusted hazard ratio was 1.16 overall and 1.50 for tumors of the central nervous system. Hazard ratios (interpreted like relative risks) of this size are more likely due to bias than to causation (Shapiro, 2000). Only a properly done randomized controlled trial free of bias could discern effect sizes this small (Schulz and Grimes, 2006).

Pseudo-epidemics in administrative databases

Other Scandinavian studies linking administrative databases have recently caused false alarms. Several reports on medical abortion have used the Finnish Abortion Registry (Niinimäki *et al.*, 2009, 2011). The reported adverse event rates were nonsensical: 20 to 100 times higher than rates reported by others (Grimes and Raymond, 2011). When



challenged in a letter to the editor (Fjerstad *et al.*, 2010), the Finnish authors acknowledged that, 'many of the "complications" are not really such, but rather concerns... that bring women back to the health care system.' The reports had counted office visits by the 'worried well' as adverse events, which is misleading.

In Denmark, studies of venous thromboembolism using the National Registry of Patients have produced red herrings as well (Grimes, 2010). For example, reports in high-profile general medical journals claimed that the levonorgestrel intrauterine system was associated with statistically significant protection against venous thrombosis (Lidegaard *et al.*, 2012b) and thrombotic stroke (Lidegaard *et al.*, 2012a). While this intra-uterine device (IUD) has several established non-contraceptive health benefits (Fraser, 2013), prevention of clots and prevention of stroke are not among them. Without validation of outcomes (Severinsen *et al.*, 2010; U.S. Food and Drug Administration) and adequate control for potential confounding (Dinger and Shapiro, 2012), these database studies (Lidegaard *et al.*, 2012a,b) are not credible.

Epidemiologic hubris

Epidemiologists need to approach their work with greater humility and circumspection (Boffetta *et al.*, 2008). Few seem to understand that most published research findings are false (Ioannidis, 2005). Today, any result coming from an observational study is most likely wrong (Young and Karr, 2011). Among the reported associations that are real, most are exaggerated (Ioannidis, 2008). Few reports, including Reigstad *et al.* (2015a) adequately explain their weaknesses (Shapiro, 2008). Sadly, these flawed manuscripts usually get published somewhere (Smith, 2006; Ioannidis *et al.*, 2010). As noted by Sir Iain Chalmers, a great deal of rubbish is published in medical journals these days (Chalmers, 2007).

Damage from false alarms

Modern epidemiology has a checkered history (Taubes, 1995). Among its many false alarms have been purported links between cigarettes and suicide, reserpine and breast cancer, menopausal estrogen and reduced cardiovascular disease, coffee and pancreatic cancer, beta-carotene and reduced lung cancer, IUD use and infertility, and oral contraceptives and pituitary adenoma (Grimes and Schulz, 2012). The claim that ART is related to an increased risk of brain tumors (Reigstad *et al.*, 2015a) and breast cancer (Reigstad *et al.*, 2015b) can be added to the list.

Infertile couples face many challenges: social, emotional and financial (Cousineau and Domar, 2007; Chambers *et al.*, 2013). Compounding their suffering by frightening them about brain tumors (Reigstad *et al.*, 2015a) and breast cancer (Reigstad *et al.*, 2015b) is cruel. Newspapers love sensational medical claims, no matter how implausible (Russell, 1999). These media scares hurt women (Jones *et al.*, 1980). For example, the 1995 media blitz about the alleged (and subsequently refuted) hazards of certain oral contraceptives (Dinger *et al.*, 2007, 2014) was temporally associated with a preventable epidemic of unplanned pregnancies and abortions (Skjeldestad, 1997).

Research priorities

A finite amount of money and resources can be spent on women's health research. Hence, no research project can be considered in a vacuum. One must always ask what benefits would have accrued had the same

amount of money and resources been devoted to other health problems. If the other problems have greater importance, then the research in question has a net negative effect on public health. Linking and dredging administrative databases in search of a problem (rather than addressing a real one) wastes resources and is thus unethical (*American College of Obstetricians and Gynecologists*, 2004).

The way ahead

Medicine today needs less research, better research and research done for the right reasons (Altman, 1994). The incentives, however, are still wrong; researchers are rewarded for the quantity and not the quality of their publications. Medicine also needs 'better training and more circumspection on the part of investigators, tougher editorial standards on the part of journals, and hefty skepticism on the part of referees and readers' (Grimes and Schulz, 2012). As a Harvard epidemiologist ruefully noted, 'We [epidemiologists] are fast becoming a nuisance to society. People don't take us seriously anymore, and when they do take us seriously, we may unintentionally do more harm than good.' (Taubes, 1995).

Author's roles

D.A.G. wrote, reviewed, critically revised and approved the final manuscript.

Funding

No external funding was either sought or obtained for this study.

Conflict of interest

D.A.G. serves on several Scientific Monitoring and Advisory Committees for studies sponsored by Bayer Pharma AG.

References

- Altman DG. The scandal of poor medical research. *BMJ* 1994;**308**:283–284.
- American College of Obstetricians and Gynecologists. *Ethics in obstetrics and gynecology*, 2nd edn. Washington, DC: American College of Obstetricians and Gynecologists, 2004.
- Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst* 2008;**100**:988–995.
- Chalmers I. Raising the standards of clinical trials and research. *Bull World Health Organ* 2007;**85**:658–659.
- Chambers GM, Adamson GD, Eijkemans MJ. Acceptable cost for the patient and society. *Fertil Steril* 2013;**100**:319–327.
- Cousineau TM, Domar AD. Psychological impact of infertility. *Best Pract Res Clin Obstet Gynaecol* 2007;**21**:293–308.
- Dinger J, Shapiro S. Combined oral contraceptives, venous thromboembolism, and the problem of interpreting large but incomplete datasets. *J Fam Plann Reprod Health Care* 2012;**38**:2–6.
- Dinger JC, Heinemann LA, Kuhl-Habich D. The safety of a drospirenone-containing oral contraceptive: final results from the European Active Surveillance Study on oral contraceptives based on 142,475 women-years of observation. *Contraception* 2007;**75**:344–354.
- Dinger J, Bardenheuer K, Heinemann K. Cardiovascular and general safety of a 24-day regimen of drospirenone-containing combined oral contraceptives: final results from the International Active Surveillance Study of Women Taking Oral Contraceptives. *Contraception* 2014;**89**:253–263.
- Fjerstad M, Westhoff C, Lifford KL. Immediate complications after medical compared with surgical termination of pregnancy. *Obstet Gynecol* 2010;**115**:660; author reply 660–661.
- Fraser IS. Added health benefits of the levonorgestrel contraceptive intrauterine system and other hormonal contraceptive delivery systems. *Contraception* 2013;**87**:273–279.
- Grimes DA. Epidemiologic research using administrative databases: garbage in, garbage out. *Obstet Gynecol* 2010;**116**:1018–1019.
- Grimes DA, Raymond EG. Medical abortion for adolescents. *BMJ* 2011;**342**:d2185.
- Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. *Lancet* 2002;**359**:341–345.
- Grimes DA, Schulz KF. False alarms and pseudo-epidemics: the limitations of observational epidemiology. *Obstet Gynecol* 2012;**120**:920–927.
- Hill AB. The environment and disease: association or causation. *Proc R Soc Med* 1965;**58**:295–300.
- Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;**2**:e124.
- Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;**19**:640–648.
- Ioannidis JP, Tatsioni A, Karassa FB. Who is afraid of reviewers' comments? Or, why anything can be published and anything can be cited. *Eur J Clin Invest* 2010;**40**:285–287.
- Jones EF, Beniger JR, Westoff CF. Pill and IUD discontinuation in the United States, 1970–1975: the influence of the media. *Fam Plann Perspect* 1980;**12**:293–300.
- Lam JU, Rebolj M, Dugue PA, Bonde J, von Euler-Chelpin M, Lyng E. Condom use in prevention of Human Papillomavirus infections and cervical neoplasia: systematic review of longitudinal studies. *J Med Screen* 2014;**21**:38–50.
- Levine MA, El-Nahas AI, Asa B. Relative risk and odds ratio data are still portrayed with inappropriate scales in the medical literature. *J Clin Epidemiol* 2010;**63**:1045–1047.
- Lidegaard O, Lokkegaard E, Jensen A, Skovlund CW, Keiding N. Thrombotic stroke and myocardial infarction with hormonal contraception. *N Engl J Med* 2012a;**366**:2257–2266.
- Lidegaard O, Nielsen LH, Skovlund CW, Lokkegaard E. Venous thrombosis in users of non-oral hormonal contraception: follow-up study, Denmark 2001–10. *BMJ* 2012b;**344**:e2990.
- Ness RB, Grisso JA, Klapper J, Schlesselman JJ, Silberzweig S, Vergona R, Morgan M, Wheeler JE. Risk of ovarian cancer in relation to estrogen and progestin dose and use characteristics of oral contraceptives. SHARE Study Group. Steroid Hormones and Reproductions. *Am J Epidemiol* 2000;**152**:233–241.
- Nevadunsky NS, Van Arsdale A, Strickler HD, Moadel A, Kaur G, Levitt J, Girda E, Goldfinger M, Goldberg GL, Einstein MH. Obesity and age at diagnosis of endometrial cancer. *Obstet Gynecol* 2014;**124**:300–306.
- Niinimäki M, Pouta A, Bloigu A, Gissler M, Hemminki E, Suhonen S, Heikinheimo O. Immediate complications after medical compared with surgical termination of pregnancy. *Obstet Gynecol* 2009;**114**:795–804.
- Niinimäki M, Suhonen S, Mentula M, Hemminki E, Heikinheimo O, Gissler M. Comparison of rates of adverse events in adolescent and adult women undergoing medical abortion: population register based study. *BMJ* 2011;**342**:d2111.
- Reigstad MM, Larsen IK, Myklebust TÅ, Røsbjerg TE, Oldereid NB, Omland AK, Vangen S, Brinton LA, Storeng R. Cancer risk among parous women following assisted reproductive technology. *Hum Reprod* 2015a;**30**:1952–1963.
- Reigstad MM, Larsen IK, Myklebust TÅ, Røsbjerg TE, Oldereid NB, Omland AK, Vangen S, Brinton LA, Storeng R. Risk of breast cancer following fertility treatment—a registry based cohort study of parous women in Norway. *Int J Cancer* 2015b;**136**:1140–1148.

- Russell C. Living can be hazardous to your health: how the news media cover cancer risks. *J Natl Cancer Inst Monogr* 1999;167–170.
- Schulz KF, Grimes DA. *The Lancet Handbook of Essential Concepts in Clinical Research*. London: Elsevier, 2006
- Severinsen MT, Kristensen SR, Overvad K, Dethlefsen C, Tjønneland A, Johnsen SP. Venous thromboembolism discharge diagnoses in the Danish National Patient Registry should be used with caution. *J Clin Epidemiol* 2010;63:223–228.
- Shapiro S. Bias in the evaluation of low-magnitude associations: an empirical perspective. *Am J Epidemiol* 2000;151:939–945.
- Shapiro S. Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy Part 2. Principles of causality in epidemiological research: confounding, effect modification and strength of association. *J Fam Plann Reprod Health Care* 2008;34:185–190.
- Silver N. *The Signal and the Noise*. New York: The Penguin Press, 2012.
- Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24.
- Skjeldestad FE. Increased number of induced abortions in Norway after media coverage of adverse vascular events from the use of third-generation oral contraceptives. *Contraception* 1997;55:11–14.
- Smith GD. Reflections on the limitations to epidemiology. *J Clin Epidemiol* 2001;54:325–331.
- Smith R. The highly profitable but unethical business of publishing medical research. *J R Soc Med* 2006;99:452–456.
- Taubes G. Epidemiology faces its limits. *Science* 1995;269:164–169.
- U.S. Food and Drug Administration. Best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data sets. <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm> (accessed May 10, 2015).
- von Elm E, Egger M. The scandal of poor epidemiological research. *BMJ* 2004;329:868–869.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370:1453–1457.
- Wakeham K, Kavanagh K. The burden of HPV-associated anogenital cancers. *Curr Oncol Rep* 2014;16:402.
- Winkelstein WJ. Smoking and cervical cancer—current status: a review. *Am J Epidemiol* 1990;131:945–957; discussion 958–960.
- Young SS, Karr A. Deming, data and observational studies. A process out of control and needing fixing. *Significance* 2011:120.