What Can We Learn From Large Databases? Lessons From Autoimmunity

Howard Amital MD MHA

Department of Medicine B, Sheba Medical Center, Tel Hashomer, affiliated with Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

ABSTRACT: The increasing use of computerized medical records has made the clinical data of the entire population available for epidemiological research. The resultant accessibility to this information mandates careful adaptions of ethical guidelines regarding the handling of clinical data. At the same time it grants a unique opportunity to explore the clinical nature of health and disease in large populations across all of society's strata, socioeconomic levels, ethnicities, and geographic locations regardless of their vicinity or distance to tertiary care centers. Analysis of large databases allows us to learn the public's behavior towards medical services and to investigate how medical interventions affect outcomes over time. Moreover, the interaction between different co-morbidities can be better understood by large population studies. The huge numbers of patients involved in these studies provide a good model of multivariate analysis, a statistical tool that by following proper population adjustments underlines the true independent associations between different conditions. Nevertheless, the limitations of these studies should be borne in mind, such as in-built imprecision of diagnoses, incompleteness of the medical data, and the fact that these databases were initially planned for clinical and not investigational use.

IMAJ 2016; 18: 225–227

KEY WORDS: medical records, randomized controlled studies, meta-analysis, case series, databases

The transformation of medical records from a paper to a digital medium has contributed significantly to quality assurance, prevented the loss of medical files, reduced errors such as prescription mistakes, and increased the accessibility to care providers regardless of previous geographic and organizational barriers.

In addition to quality improvement, computerized medical records enable medical research in ways previously not imagined. With the availability of the entire population's demographic and clinical data, the possibilities of studying human diseases are infinite. At the same time of course, this mandates modifying the ethical and moral approaches we used when handling data and conducting research [1-6].

In this short paper, using several examples, I will present the advantages of "big data" analysis derived from clinical records of the largest health maintenance organizations in Israel compared to the commonly used methodologies we accepted as optimal from our earliest years in the medical profession.

PITFALLS OF COMMONLY USED METHODOLOGIES

Although randomized trials are the most robust critical method, poorly designed studies are susceptible to different forms of bias. A randomized double-blinded study is considered the ultimate desired method for evaluating any medical intervention; it is believed to be the optimal assessment tool for any medical intervention, whether a new medication, a surgical procedure, or a novel device that has been developed. Nevertheless, despite scrupulous study design, the interests of the pharma industry often obfuscate the basic concepts behind proper medical research. Such approaches may create prevailing medical concepts, eventually biasing clinical studies toward achievements that result in more tangible and profitable study outcomes.

To illustrate I will use an example in my area of specialization, rheumatology. I have deliberately chosen a well-designed study that was published in one of the leading medical journals, the *Lancet*. The Adacta study drew much attention several years ago when it overturned barriers that many pharmaceutical companies had erected to evade the need for direct head-to-head comparisons between different biological agents. Such hesitations evolved over the years due to the heavy economic implications of such costly studies and to the concern that a certain product will not be found superior to the comparator. In this study two excellent biological agents commonly used as part of the therapeutic armamentarium for patients with rheumatoid arthritis (RA) were compared [7]. Both medications, tocilizumab (Actemra®, manufactured by Roche) and adalimumab (Humira®, manufactured by AbbVie) were shown in previous studies to provide clinical and radiological efficacy [8,9]. However, tocilizumab was found to have a significant suppressive effect on acute-phase reactants such as C-reactive protein (CRP) and erythrocyte sedimentation rate, both routinely used as significant components of the composite primary outcome of the study (DAS-28) [7]. Moreover, studies prior to the Adacta demonstrated that tocilizumab is different from other biologic agents due to its relative independence of concomitant metho-

trexate use. In comparing tocilizumab to adalimumab, head to head without the important contribution of the background methotrexate, the results were determined in advance prior to the beginning of the study, as were the outcomes.

Another method whereby prospective clinical studies affect outcomes is by generating selective patient groups in a way that will not impair the results. Safety is often the major concern of the pharma industry. Inclusion and exclusion criteria usually do not attract the attention of the medical readership, yet these parameters often have a strong impact on the study outcomes, rendering them more favorable when certain individuals are not enrolled in the study. Such measures obviously bias the results, making them less applicable to the general patient population that usually does not comply with all the stringent study recruitment definitions [10]. In their analysis of psychiatric outpatients with major depressive disorder, Zimmerman et al. [11] found that randomized controlled studies would have excluded 86% of them due to their having co-morbid anxiety or substance use disorders, insufficient depressive symptoms, and/or current suicidal ideation.

Another methodology that is currently much used is the meta-analysis. Although the main advantage of this method is the large number of subjects included in the final report, meta-analyses have inborn faults. Pooled analyses should only be performed when study methodologies share similarities. Since many studies use dissimilar parameters such as different inclusion and exclusion criteria, duration of treatment and primary outcomes, the data cannot be easily pooled. Often the demographics of the subjects in the study differ in age, ethnicity, and disease duration and severity, making the data outcomes questionable and uninterpretable. Almost always, meta-analyses researchers suffice with the published data since pharmaceutical companies are reluctant to provide free accessibility to their original crude data. Above all, positive results are published more frequently, and studies with less favorable outcomes ae omitted from the final analysis.

Case series and case reports have mainly educational importance; they are often prepared by young students and physicians at the dawn of their academic career. These publications present interesting clinical material which by nature is biased; the authors' contribution is to underline interesting clinical findings or promising interventions. In these publications there is no concern or intention to control these observations. Of course, parallel negative findings will not find their way to print.

THE ADVANTAGES OF ANALYSIS OF "REAL-LIFE DATA"

Using real-life data provides multiple assets to medical research. In contrast to the randomized controlled studies, there is no need to select a "representative" sample if the entire population can be investigated. The validity of real-life data versus small cohorts is reflected in the following example. Novack and co-authors [12] reported a strong predominance of males (63/91) among offspring born to 42 mothers with systemic lupus erythematosus (SLE), a finding perhaps indicating that the female genotypic expression itself elicits an immune response, which leads to premature termination in utero of the development of female fetuses. To further challenge this intriguing observation, we studied the population of all females aged 16 to 46 who were members of the Maccabi Health Services during the period 2000–2011 and had at least one pregnancy. From a total of 182,073 women who had at least one indication of pregnancy during the study period, 270 were diagnosed with SLE. The proportion (95% confidence intervals) of male offspring born to lupus mothers in our study was 51.8% (95% confidence interval 46.6–57.0%), rendering Novack's interesting hypothesis invalid [13].

Due to the concurrent availability of logistic information, such drug-dispensing real-life data analysis may also shed light on the medical behavior of the population, such as adherence and persistence, and on duration of therapy. An example for such a study evolved from an enigma that troubled me for years, why patients with gout develop recurrent acute gouty attacks despite effective therapy with urate-lowering medications? In our study we analyzed data derived from the Maccabi Health Services database regarding patients with the diagnosis of gout who were treated with allopurinol over a 7 year period (2002–2008). They were all assessed for their degree of adherence and persistence regarding allopurinol prescriptions provided by their family physicians. A total of 7644 patients were identified, predominantly men (72%). Interestingly, only 1331 (17%) of these patients were compliant with allopurinol therapy, 36% complied partially, and 47% complied poorly. The average time to discontinuation of therapy in men compared to women was 358 and 379 days respectively. By using logistic regression we identified a high risk group for non-compliance, namely, women aged 45–65 years. In contrast, compliance was achieved among those with chronic illnesses, particularly cardiovascular disease. Such data have much importance for directing educational and financial resources in order to improve poor clinical outcomes. This information is of particular interest since it reflects the socioeconomic strata, ethnicities and geographic distribution of the entire Israeli population [14].

Since medical computerized records incorporate all the clinical and logistic information of patients over years, they can be utilized for a comprehensive analysis of certain interventions on disease development years before its emergence. An example of such an attempt is demonstrated in our study [17] showing the anti-inflammatory potential of statins in preventing the future development of RA. Statins have only a modest beneficial effect in patients with RA when added to standard therapy. This observation was reported in small randomized clinical trials showing only a marginal beneficial effect on disease activity and CRP levels [15,16]. In our study [17] we used the computerized medical databases of the Maccabi Health Services database and identified RA cases among adults who began statin therapy between

1998 and 2007. A total of 211,627 individuals were eligible for the RA cohort analysis. During the study follow-up there were 2578 incident RA cases (3.07 per 1000 person-years). The crude incidence density rate of RA among non-adherent patients was 51% higher compared to patients who were covered with statins for at least 80% of the follow-up period. Larger differences were observed in younger patients and in patients initiating treatment with high efficacy statins [17]. These data demonstrated the potency of large-scale studies to underline different aspects of medical interventions carried over years, a format almost impossible for other methods of analysis.

An additional advantage of database analysis is its validity in studying coexisting co-morbidities, since covering all population subgroups provides an accurate assessment of such issues. Since these patients are often exempted from randomized clinical studies, no other tools can reliably provide such an understanding of disease associations. In recent studies we have been able to decipher such linkages with data from Clalit Health Services (the largest of the four health funds in Israel). We found coexistence between SLE and hyperthyroidism, rheumatoid arthritis and chronic obstructive lung disease, and RA and bipolar disease [18-20].

One cannot deny the drawbacks of this methodology; the data analysis is not always performed initially for research purposes; the patients are classified by experts; and classification criteria and the data are often incomplete, sometimes lacking important elements. Nevertheless, given the huge number of individuals included in such studies there is much to be learned from "real-life" data analysis. Many national databases are changing our comprehension of disease and therapy and it is anticipated that they will continue to do so.

Correspondence

Dr. H. Amital

Head, Dept. of Internal Medicine B, Sheba Medical Center, Tel Hashomer 526521, Israel

email: howard.amital@sheba.health.gov.il

References

1. Arnaud L, Fagot JP, Mathian A, Paita M, Fagot-Campagna A, Amoura Z. Prevalence and incidence of systemic lupus erythematosus in France: a 2010 nation-wide population-based study. *Autoimmun Rev* 2014; 13 (11): 1082-9.
2. Porcelli B, Pozza A, Bizzaro N, et al. Association between stressful life events and autoimmune diseases: a systematic review and meta-analysis of retrospective case-control studies. *Autoimmun Rev* 2016; 15: 325-34.
3. Kokia ES, Marom R, Shalev V, Jan Y, Shemer J. The use of medical informatics as a management tool for community health services during the 2006 Israel-Lebanon War. *IMAJ* 2006; 8 (12): 865-9.
4. Milton CL. Information sharing: transparency, nursing ethics, and practice implications with electronic medical records. *Nurs Sci Q* 2009; 22 (3): 214-19.
5. Bar-On Y, Shalev V, Weitzman D, Chodick G, Amital H. Old obstacles but new hopes: trying to understand the fibromyalgia construct. *IMAJ* 2014; 16 (10): 625-6.
6. Yaniv G, Twig G, Shor DB, et al. A volcanic explosion of autoantibodies in systemic lupus erythematosus: a diversity of 180 different antibodies found in SLE patients. *Autoimmun Rev* 2015; 14 (1): 75-9.
7. Gabay C, Emery P, van Vollenhoven R, Dikranian A, et al. Tocilizumab monotherapy versus adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA): a randomised, double-blind, controlled phase 4 trial. *Lancet* 2013; 381 (9877): 1541-50.
8. Keystone EC, Kavanaugh AF, Sharp JT, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum* 2004; 50 (5): 1400-11.
9. Nishimoto N, Hashimoto J, Miyasaka N, et al. Study of active controlled monotherapy used for rheumatoid arthritis, an IL-6 inhibitor (SAMURAI): evidence of clinical and radiographic benefit from an x ray reader-blinded randomised controlled trial of tocilizumab. *Ann Rheum Dis* 2007; 66 (9): 1162-7.
10. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* 2003; 327 (7418): 785-9.
11. Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am J Psychiatry* 2002; 159 (3): 469-73.
12. Novack V, Erez O, Novack L, Jotkowitz A, Meir A, Mazor M. Sex distribution of newborns to mothers with systemic lupus erythematosus. *Epidemiology* 2006; 17 (3): 341-2.
13. Dar L, Shalev V, Weitzman D, Chodick G, Arnson Y, Amital H. No male predominance in offspring of women with rheumatoid arthritis or systemic lupus erythematosus. *Immunol Res* 2014; 60 (2-3): 361-5.
14. Zandman-Goddard G, Amital H, Shamrayevsky N, Raz R, Shalev V, Chodick G. Rates of adherence and persistence with allopurinol therapy among gout patients in Israel. *Rheumatology (Oxford)* 2013; 52 (6): 1126-31.
15. McCarey DW, McInnes IB, Madhok R, et al. Trial of atorvastatin in Rheumatoid Arthritis (TARA): double-blind, randomised placebo-controlled trial. *Lancet* 2004; 363 (9426): 2015-21.
16. Okamoto H, Koizumi K, Kamitsuji S, et al. Beneficial action of statins in patients with rheumatoid arthritis in a large observational cohort. *J Rheumatol* 2007; 34 (5): 964-8.
17. Chodick G, Amital H, Shalem Y, et al. Persistence with statins and onset of rheumatoid arthritis: a population-based cohort study. *PLoS Med* 2010; 7 (9): e1000336.
18. Bieber V, Cohen AD, Freud T, Agmon-Levin N, Gertel S, Amital H. Autoimmune smoke and fire – coexisting rheumatoid arthritis and chronic obstructive pulmonary disease: a cross-sectional analysis. *Immunol Res* 2013; 56 (2-3): 261-6.
19. Farhi A, Cohen AD, Shovman O, Comaneshter D, Amital H, Amital D. Bipolar disorder associated with rheumatoid arthritis: a case-control study. *J Affect Disord* 2016; 189: 287-9.
20. Watad A, Cohen AD, Comaneshter D, Tekes-Manova D, Amital H. Hyperthyroidism association with SLE, lessons from real-life data – a case-control study. *Autoimmunity* 2016; 49: 17-20.

"The ideals which have lighted my way, and time after time have given me new courage to face life cheerfully, have been Kindness, Beauty, and Truth"

Albert Einstein (1879-1955), German theoretical physicist and Nobel laureate

"I worked my way up from nothing to a state of extreme poverty"

Groucho Marx (1890-1977), American comedian and film and television star, known as a master of quick wit and regarded widely as one of the best comedians of the modern era