CellPress

Data collection as a barrier to personalized medicine

David Zakim[1] and Matthias Schwab[2,3]

[1] Institute for Digital Medicine Foundation, 70192 Stuttgart, Germany
[2] Dr Margarete Fischer-Bosch-Institute of Clinical Pharmacology, 0376 Stuttgart, Germany
[3] Department of Clinical Pharmacology, University Hospital Tübingen, 72076 Tübingen, Germany

Basic life science research holds the promise of personalizing medical care. However, translation steps from the laboratory to the bedside are not trivial. Results from clinical research are difficult to replicate in part because study cohorts are poorly defined phenotypically. Here, we discuss how computer technology can improve the collection of clinical data to enable translation of insights from basic science to validated clinical guidelines.

Barriers to personalized medicine

Standardized methodology for collecting all data, from all samples, that can affect the outcome of an experiment is a bedrock principle of experimental science. Disregarding this principle, which also applies to clinical research, can lead to nonreplicable results. Indeed, an absence of standardized methodology for collecting clinical phenotypes is a problem for clinical research [1,2] and a barrier to translating advances in basic medical science to validated guidelines. Here, we focus attention on the data collection issues that limit the translation of advances in basic medical science to useful clinical guidelines. We use the example of studies on the efficacy of clopidogrel to illustrate how data problems lead to clinical research with limited value. We propose a general solution to the problem of standardizing data collection in clinical research protocols while extending the data collected to all elements that could affect outcomes. We also propose a new paradigm for the conduct of clinical research not only to generate reproducible results, but also to extend the range of problems that can be investigated through reduced costs of research.

Clinical research is limited by deficient data collection

Ischemic heart disease is the number-one cause of death worldwide, for which the lesion is an atheromatous plaque in a coronary artery. When ruptured or eroded, platelets bind to the plaque, aggregate, and then interact with the clotting system to form an occluding thrombus. Clopidogrel inhibits the aggregation step in this process by binding at the ADP-binding site on platelet membranes, which is its putative therapeutic mechanism [3]. Clopidogrel plus aspirin, which binds to platelets at a site that is different from the ADP receptor, is more effective [4]. Yet, this combination fails to prevent ischemic events in most treated patients, and both drugs cause bleeding. Given that clopidogrel is a prodrug oxidized to the active metabolite by cytochrome P450s and that the most significant of these enzymes (CYP2C19) [5] has multiple loss-of-function alleles (http://www.cypalleles.ki.se/cyp2c19.htm), it was proposed that the *CYP2C19* genotype would predict efficacy and enable individualized dosing to maximize efficacy while minimizing adverse effects. However, clinical testing of these ideas yielded nonreplicable results [6,7]. A meta-analysis of all data found only a marginal association in one study [8], whereas another meta-analysis including only patients with acute coronary syndromes provided strong evidence of CYP2C19 variants on outcomes [9]. Further confounding the issue are nonreproducible results for the effect of proton pump inhibitors (PPIs) on the efficacy of clopidogrel despite their inhibition of the CYP2C19-catalyzed oxidation of clopidogrel [10]. Two important questions need to be addressed in the context of these results: how do we account for the nonreproducibility of results; and what explains the absence of an effect of metabolism on efficacy?

Figure 1 illustrates the complexity of the pathology of recurrent ischemic events: genetics affecting metabolism of clopidogrel; interaction of clopidogrel with platelets; and risks for coronary events, including risks from medications such nonsteroid anti-inflammatory drugs [11]. Drugs prolonging the QT interval and heart failure are listed because arrhythmias may cause sudden death (monitored in clopidogrel trials) in the absence of a coronary event. Obviously, there is extensive heterogeneity of patients with coronary disease in clopidogrel trials. There is also heterogeneity in the criteria for outcomes. Failure to control data collection to include all factors impacting these variables could explain the nonreproducible results from the clopidogrel work. There is no way to verify, for example, that different studies of the *CYP2C19* genotype in relation to efficacy of clopidogrel were repeat experiments (i.e., that the same cohorts were studied in all trials or that the same outcomes were measured). We note that the true failure to find a relation between metabolism and efficacy implies that the putative therapeutic mechanism of clopidogrel is incorrect.

The 'clopidogrel problem' is not unique [12,13]. However, analysis of the heterogeneity and complexity of the clinical problem of ischemic disease on the one hand and the

Corresponding author: Zakim, D. (dzakim@sonic.net).
Keywords: personalized medicine; CLEOS; computerized history-taking program.

0165-6147/
 http://dx.doi.org/10.1016/j.tips.2014.11.002

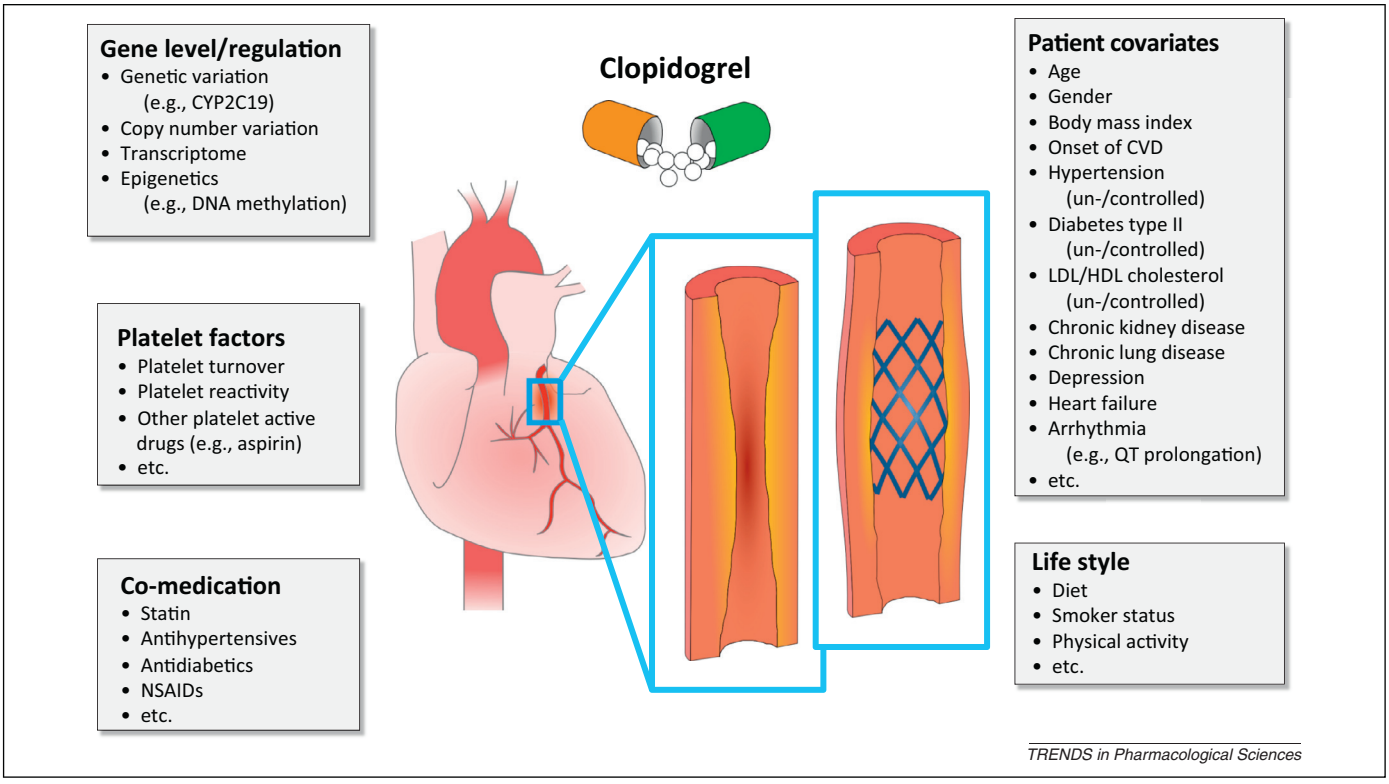

Figure 1. Selection of attributes to 'control' to establish a relation between the *CYP2C19* genotype and efficacy of clopidogrel. Listed comorbid states have known negative impacts on the incidence and outcomes in coronary disease. Abbreviations: CVD, cardiovascular disease; L/HDL, low-/high-density lipoproteins; NSAIDs, nonsteroidal anti-inflammatory drugs.

complexity of factors that affect drug action on the other, indicates that it is not tractable to conduct a prospective, clopidogrel trial that controls all the experimental variables in Figure 1. The definable subsets of patients require trials with millions of people. While there is no shortage of patients treated with clopidogrel or eligible for treatment, there is not enough money to mount prospective trials of this size. The complexity of many clinical issues leaves no alternative but to base clinical research on data collected from large populations of patients treated in everyday practice. Thus, the challenge for the clinical research enterprise is the development of methods for collecting standardized clinical data in routine practice [14] and not to rely on data collected by different physicians using variable funds of knowledge, cognitive skills, available time, and unverifiable heuristics.

Automated methods to collect clinical data

There is optimism that electronic medical records (EMRs) could solve the issue of data collection in everyday practice [15,16], but this literature conflates the value of computing to store data collected manually with the value of computing as a high-throughput technology for collecting standardized data. EMRs do not change the physician's fund of knowledge or use of heuristics and do not automate, scale, or standardize data collection. Data in EMRs are cut and pasted between files for different patients [17], and there is widespread fraud (https://oig.hhs.gov/oei/reports/oei-05-11-00250.asp). EMRs require significantly more time for data entry compared with paper charts [18] and reduce physician productivity unless data entry is off-loaded to other staff [19]. By contrast, control over data collection can be achieved with software to emulate the clinical thinking and decision-making of expert clinicians acquiring medical histories. Computers programmed in this way can acquire detailed clinical phenotypes by standardized protocol. This use of computing can remove limitations on the completeness of data imposed by time constraints and variable funds of physician knowledge and can standardize data collection for all patients. This makes it possible to build standardized, detailed, longitudinal clinical databases covering all patients in contact with a healthcare system to empower clinical research and remove the lack of phenotype as an impediment to progress in -omics research. On the clinical side, this technology leverages the expertise of leading clinicians worldwide and, when coupled to decision support, delivers relevant knowledge to the point of care.

Description of a potential solution

There is a long history of interest in software programs for interviewing patients [20–22]. None of these was designed to empower clinical research and everyday medical practice, and none has had a sufficient knowledge base and dynamic structure to emulate clinical thinking [23]. One of us has developed a program with these properties. Clinical Expert Operating System (CLEOS®) (Figure 2) interacts with patients at a text- and graphic-driven interface, which the patient can use while at home, work, hospital, or clinic. Patients with an ID and password can access their file from any computer terminal to enter data. The process of data entry begins with the entry of demographics and then a chief complaint. The first interview proceeds with acquisition of a detailed history directed by the differential of the chief complaint. Following completion of the history of the present illness, a life-long history across all risks and organ systems