Diabetes Care Editors' Expert Forum 2018: Managing Big Data for Diabetes Research and Care

Matthew C. Riddle,[1] Lawrence Blonde,[2] Hertzel C. Gerstein,[3] Edward W. Gregg,[4] Rury R. Holman,[5] John M. Lachin,[6] Gregory A. Nichols,[7] Alexander Turchin,[8] and William T. Cefalu[9]

Technological progress in the past half century has greatly increased our ability to collect, store, and transmit vast quantities of information, giving rise to the term "big data." This term refers to very large data sets that can be analyzed to identify patterns, trends, and associations. In medicine—including diabetes care and research—big data come from three main sources: electronic medical records (EMRs), surveys and registries, and randomized controlled trials (RCTs). These systems have evolved in different ways, each with strengths and limitations. EMRs continuously accumulate information about patients and make it readily accessible but are limited by missing data or data that are not quality assured. Because EMRs vary in structure and management, comparisons of data between health systems may be difficult. Registries and surveys provide data that are consistently collected and representative of broad populations but are limited in scope and may be updated only intermittently. RCT databases excel in the specificity, completeness, and accuracy of their data, but rarely include a fully representative sample of the general population. Also, they are costly to build and seldom maintained after a trial's end. To consider these issues, and the challenges and opportunities they present, the editors of *Diabetes Care* convened a group of experts in management of diabetes-related data on 21 June 2018, in conjunction with the American Diabetes Association's 78th Scientific Sessions in Orlando, FL. This article summarizes the discussion and conclusions of that forum, offering a vision of benefits that might be realized from prospectively designed and unified data-management systems to support the collective needs of clinical, surveillance, and research activities related to diabetes.

[1]*Division of Endocrinology, Diabetes & Clinical Nutrition, Oregon Health & Science University, Portland, OR*
[2]*Ochsner Diabetes Clinical Research Unit, Frank Riddick Diabetes Institute, Department of Endocrinology, Ochsner Medical Center, New Orleans, LA*
[3]*McMaster University and Hamilton Health Sciences, Hamilton, Ontario, Canada*
[4]*Centers for Disease Control and Prevention, Atlanta, GA*
[5]*Diabetes Trial Unit, Radcliffe Department of Medicine, University of Oxford, Oxford, U.K.*
[6]*The George Washington University Biostatistics Center, Rockville, MD*
[7]*Center for Health Research, Kaiser Permanente Northwest, Portland, OR*
[8]*Brigham and Women's Hospital and Harvard Medical School, Boston, MA*
[9]*American Diabetes Association, Arlington, VA*

Corresponding author: Matthew C. Riddle, riddlem@ohsu.edu

Received 20 March 2019 and accepted 20 March 2019

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

This article is featured in a podcast available at http://www.diabetesjournals.org/content/diabetes-core-update-podcasts.

Within the span of their professional careers, older physicians and investigators have experienced a revolution in the management of data. In the 1960s, we wrote chart notes and prescriptions by hand, pored over large volumes in libraries, recorded notes from these tomes on 3- by 5-inch notecards, computed means and standard deviations on mechanical calculators, and composed manuscripts for publication on typewriters. Digital technologies ushered in a paradigm change for all of these practices. Large, slow mainframe computers were developed in that decade and, concurrently, defense and academic groups established electronic communication networks. These innovations were followed in the 1970s by smaller, yet more powerful, computers and, in the 1980s, by personal computers and expanded networks. Now we have the Internet, the World Wide Web, and "cloud" storage capabilities, all of which can be accessed anywhere and at any time by individuals with smartphones and other small electronic devices. Our ability to collect, store, analyze, and transmit data has increased remarkably, giving rise to the collective term "big data": extremely large data sets that can be analyzed to identify patterns, trends, and

associations. It is now, at least in principle, possible to manage huge quantities of data over decades of time and among regions globally.

These tools for data management have long been recognized as relevant to our efforts to improve health care (1), and certainly this applies to clinical care and research in the field of diabetes. Some notable examples deserve mention. Henry J. Kaiser, a prominent defense contractor, developed health systems for employees at his shipyards in the 1940s. The Kaiser systems applied business principles to health care, including early adoption of electronic medical records (EMRs) (2). The U.S. Department of Veterans Affairs created an electronic database for its geographically dispersed medical systems in the 1980s (3). Population-based medical registries have been established in the U.K. (4) and other countries. In the U.S., the National Health and Nutrition Examination Survey (NHANES) began in 1971, and data collection continues to the present time (5). Likewise, the first large, randomized trials testing interventions for diabetes were facilitated by digital data management. The UK Prospective Diabetes Study (UKPDS) was launched in 1977 (6), and enrollment in the Diabetes Control and Complications Trial (DCCT) began in 1983 (7).

Despite the strong influence of digital technology on these projects, the systems used for clinical care, epidemiologic surveillance, and interventional trials have grown and evolved in quite different ways. Their purposes and designs differ considerably, and collected data are not easily compared among systems. To consider these issues, and both the challenges and opportunities presented by them, *Diabetes Care* convened a group of experts in the field of diabetes digital technology on 21 June 2018. Here, we report a summary of the discussion and

conclusions of that forum. The discussion was divided into the three categories of data-management systems briefly described in Table 1.

ELECTRONIC MEDICAL RECORDS: ENORMOUS POTENTIAL BUT LIMITATIONS

Ancient Egyptian physicians recorded their patients' medical information on papyri (8), and until recently, handwritten records continued to be the norm. However, as digital technology developed, it was quickly applied to medical records. Some health systems introduced electronic data management in the 1960s, with the focus initially on scheduling and billing. Over time, electronic medical record (EMR) systems have expanded to other aspects of patient care and, in the past decade, a growing number of health care organizations have largely abandoned paper-based records.

The potential of EMRs to make patient-related information more accessible is enormous. In 1863, Florence Nightingale complained that, "In attempting to arrive at the truth, I have applied everywhere for information, but in scarcely an instance have I been able to obtain hospital records fit for any purposes of comparison" (9). Recent studies have demonstrated that use of EMRs can improve preventive health services, decrease medication errors, and facilitate population health management (10–13). In the case of diabetes, analyses of data from EMRs have been shown, in appropriate settings, to help improve success in controlling glycemia, lipids, and blood pressure and to reduce the frequency of emergency department visits and nonelective hospitalizations (14–16).

At the organizational level, review of EMR data allows for the assessment of clinical visit scheduling and reimbursement, attendance and wait times,

medication prescription and dispensation, and tracking of variously defined measures of quality of care. For providers, EMRs allow immediate access to patients' clinical histories, physical and laboratory findings, and other care-related information. Providers potentially can access clinical information independent of where they or their patients may be at a given time. The importance of this ability was illustrated by the experience after Hurricane Katrina struck New Orleans and nearby areas in 2005. Clinicians who had EMR access could provide information and advice and fill prescriptions for their patients who were widely dispersed across the country, whereas those without EMRs lost contact with patients and permanently lost their paper records to storm and flood damage. Electronic records allow many different users to access medical information simultaneously and eliminate the costs of creating and delivering hard copies of records to each clinician. Virtually instantaneous remote access by on-call clinicians, including those in emergency departments and distant institutions, can assist in timely provision of care. For the care of those with diabetes, use of EMRs facilitates tracking of relevant clinical data over time, including weight, blood pressure, A1C, lipid measurements, and medications for control of various risk factors. Because a team of providers—including physicians and advanced practice providers, diabetes educators, nutritionists, and others—is typically involved in care for people with diabetes, EMRs assist in coordinating multidisciplinary care.

There are also potential limitations to the use of EMRs. Both isolated and systematic unintended consequences have been reported (17–21). Workloads of clinical providers may be increased and their morale impaired by the need to

Table 1—Typical features of current systems for managing medical data

Feature	EMRs	Public surveys and registries	RCTs
Financial support	Health system	Government	Government, industry, or voluntary health organization
Governance	System administrators	Government employees	Academic partnership with sponsor
Population included	Enrolled in public or commercial system	National or regional	Selected for study, may be international
Time of data collection	Continuous	Periodic	Specific interval

EMRs, electronic medical records; RCTs, randomized controlled trials.

enter orders for tests, prescriptions, and consultations—tasks previously performed by other health care personnel. Because of the need to review prior encounters and enter current data in the examining room, both clinicians and patients have sometimes complained about EMRs interfering with communication during visits (22,23). Although much energy is devoted to optimizing the use of EMRs in managing the logistical aspects of care (e.g., scheduling, billing, and process-based quality assessment), medical information needed for personalized management of complex conditions such as diabetes may be less easily collected, recorded, and visualized. Whereas the consistency and accuracy of entries concerning financial or operational matters are routinely checked by specialized personnel within health systems, similar quality control is rarely attempted for clinical entries. The result is variability and inconsistency in capturing even the most crucial medical information in many cases. A notable example is the difficulty of tracking insulin doses prescribed, as well as those actually taken—especially when patients are actively self-managing their glycemic control. Another is the lack of consistency in distinguishing between type 1 (autoimmune-mediated) diabetes, type 2 diabetes, and less common forms of diabetes in EMRs.

Electronic record systems come in a bewildering variety of configurations, and they frequently evolve over time. Therefore, careful implementation procedures, including user training, are crucial to their success. Although broad principles of EMR design are well established (24,25), they are not universally followed. As a result, many systems suffer from discrepancies between software design, user needs, and clinical workflow, sometimes leading to negative perceptions of their value and reliability (Fig. 1) (26–29). Alignment of EMRs with the activities and concerns of medical providers can be improved, but in many cases this is not occurring. Business-related aspects of EMR use can also pose barriers. For example, EMR system vendors may have contractual hold-harmless clauses that limit their accountability for harm or inconvenience related to defects and malfunction. It may be unclear who is responsible for maintenance of services, and difficulties may
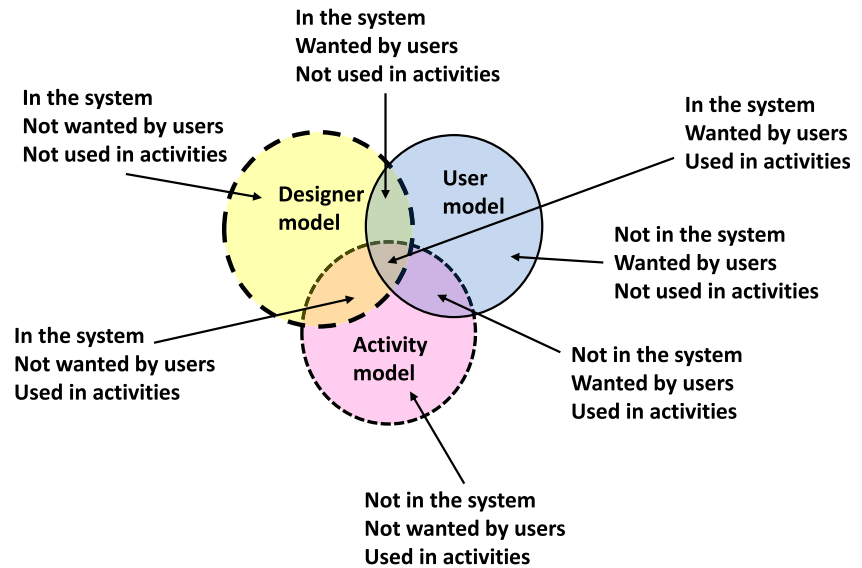
Figure 1—A conceptual model of differences between how electronic medical records are designed (Designer model), functionality desired by the users (User model), and how they are actually utilized (Activity model). Reprinted from Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011;44:1056–1067, with permission from Elsevier (29).

not be reliably reported. Governmental oversight of the quality of EMR products and services is limited (30–33).

There is considerable potential for the use of data collected routinely in EMRs for epidemiological surveillance or prospectively designed medical research (34,35). Some large health systems with long-standing databases have published useful epidemiologic reports of their experience. Notable examples relevant to diabetes include early reports of clinical inertia in advancing pharmacotherapy of diabetes (36,37) and clinical features associated with hypoglycemia in clinical practice (38). However, there are limitations to such use of data collected in EMRs under current circumstances. These include missing or unreliable data, collected without consistent definitions or quality control, and uncertain generalizability when data originate from a single institution. These problems could be addressed and some attempts have been made, although the success of such efforts depends on allocation of additional resources and support by health system administrators (39–42).

SURVEYS AND REGISTRIES: MONITORING POPULATION-WIDE TRENDS

Public health surveillance for chronic diseases has also been greatly facilitated by electronic data-management systems. Surveillance can be defined as

quantitative monitoring of population-level incidence (risk) and prevalence (frequency) of disease and of provision of preventive care, with attention to variations according to personal characteristics, time, and location (43,44). Periodic surveys can identify emerging risk factors, new health problems and comorbid conditions, gaps in care, and adverse events of treatment. Surveillance aims to identify subpopulations that are most at risk for a given disease or most likely to benefit from intervention. Data grouped according to specific characteristics of individuals may be described as a registry, which can be systematically updated to provide targeted surveillance of individuals sharing this characteristic.

Such information provides timely guidance for short-term decisions by policy makers, health plan administrators, clinicians, and the public. It also permits more in-depth etiological analyses, cost-effectiveness determinations, and health impact modeling, all relevant to long-term decisions. When combined with related disciplines (e.g., clinical epidemiology, health services and policy research, health economics, and program management evaluations), population surveillance forms the basis for public health strategies and resource allocation.

Population-level surveillance for diabetes is undergoing a rapid transformation due to new health-related data

sources and also computing and analytic approaches to large data sets (45). Diabetes surveillance in countries such as the U.S., Canada, Europe, Australia, Israel, and some Asian countries originated mainly from public survey– and direct registry–based systems. In some settings, it is now extending to include health system–based electronic registries linking EMR data, hospital and ambulatory services, laboratory and pharmacy data, and, most recently, various non-health-related data sources (46,47).

Surveillance Through Public Systems

Nationally representative surveys in the U.S. that include assessment of diabetes prevalence have existed for more than 50 years (Fig. 2), beginning with the National Health Survey in the 1960s. Next came the National Health Interview Survey (NHIS), the first National Health and Nutrition Examination Survey (NHANES I) in the 1970s, NHANES II in the 1970s and 1980s, NHANES III in 1988–1994, and continuous NHANES surveys from 1999 to the present (48–51). These are coordinated by the Centers for Disease Control and Prevention's National Center for Health Statistics.

A suite of other health care surveys—including the National Ambulatory Medical Care Survey (52), Medical Expenditure Panel Surveys from health care settings (53), and the National Hospital Discharge Survey (later supplanted by the National Inpatient Sample [NIS] [54])—collects data at the level of hospitals rather than individuals. Since 1993 the Behavior Risk Factor Surveillance System (BRFSS) has provided population-based surveys conducted at the state level (46). These surveys are complemented by registries for selected conditions such as the United States Renal Data System for end-stage renal disease (55), or for special problems and populations (e.g., the prevalence of type 1 vs. type 2 diabetes in children in the SEARCH for Diabetes in Youth study) (56). Similar evolution of surveillance has occurred in other countries as well. For example, the National Diabetes Audit in the U.K. is one of the largest annual clinical audits in the world. It integrates data from both primary and secondary care sources, with providers legally required to supply the data from their clinical practices (57).

Most of these surveys are designed to obtain repeated cross-sectional, complex samples with analytic weighting so that the estimates derived are representative of the noninstitutionalized population, including people without health insurance. NHANES is the most comprehensive survey in the U.S., consisting of a questionnaire, physical exam, and laboratory examinations every 2 years. It is the primary source for tracking total prevalence of diabetes, prediabetes, and undiagnosed diabetes, as well as selected risk factors and complications, including peripheral arterial disease, retinopathy, and chronic kidney disease (58–61). NHIS includes the single largest sample of the U.S. population and is the primary source of self-reported incidence of diagnosed diabetes. It serves as the key platform for supplemental surveys of issues ranging from health care access to preventive care (62,63). NIS is the main source of data for hospitalizations and procedures and is used to estimate and track the incidence of cardiovascular disease, stroke, and amputation (64). BRFSS has been crucial in providing state-level and, with assistance of small-area statistical modeling, county-level prevalence and incidence rates of diabetes and prevalence of obesity and physical inactivity (65). Several surveys, including NHIS and NHANES, also have linkage to the National Death Index. This is an important association that allows mortality rates to be estimated for consecutive cohorts (66). Collectively, the publicly available surveys permit researchers and policy makers to monitor a broad range of metrics such as behavioral and biochemical risk factors, preventive behaviors, receipt of preventive care, risk factor management, diabetes-related complications, disability, and mortality (Fig. 2) (43).

However, these public surveys have some fundamental limitations. First, they are largely cross-sectional data sets. Apart from the mortality linkage, the lack of longitudinal data limits assessment of changes in risk and care and the ability to examine the effectiveness of treatments or the etiology of conditions in individuals. Second, the ability to examine geographic variation in risk, care, or outcomes is limited in most surveys. Thus, their utility for directly targeting interventions to areas of greatest need in regions below the national level is impaired. While BRFSS has been useful for estimation of state- and county-level prevalence of diabetes, obesity, and physical inactivity, there are limitations of its design and data collection that allow incidence rates to be tracked reliably only at the national level (46,62,63,65). Third, although they are designed and weighted to be representative of noninstitutionalized populations, steadily declining response rates are a growing threat to validity. Finally, despite improvements in the timeliness of collection and disclosure of data, periodic surveys do not always allow real-time assessment of emerging problems. Also, incorporating new elements into the surveys requires administrative review and approval, which can be a lengthy process.

Surveillance Within Health Systems

As noted earlier, integrated health systems in the U.S. and elsewhere have used EMRs and other systematically collected data for surveillance, development of
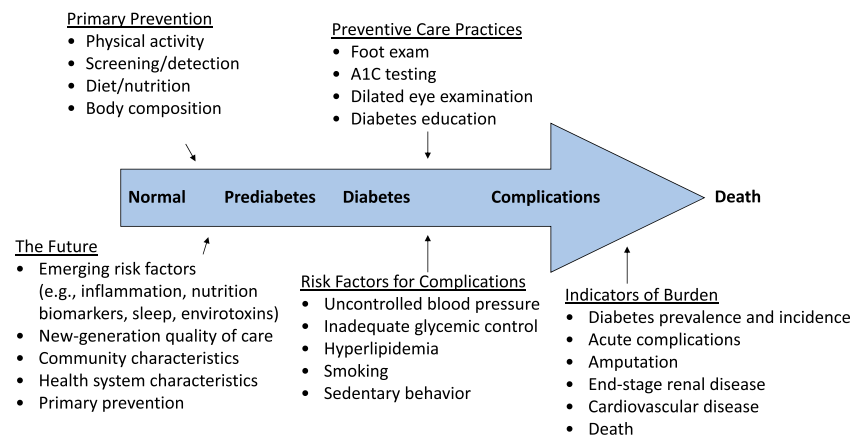
Figure 2—Overview of diabetes-related metrics monitored in the U.S. via publicly available survey data throughout the natural history of the disease. Adapted from Desai et al. (43).

registries, and evaluation of care within their populations. Direct clinical data in such systems can be linked to pharmacy and laboratory information, allowing broader assessment of processes and outcomes (67,68). This experience has set the stage for linkage of previously existing public surveys and registries to data derived from direct patient contact within private systems. This trend has been paralleled by conceptually similar population-wide registries in countries with single-payer health systems, including Sweden, Finland, Denmark, and the U.K. (47,57,69–71). Combining these registries has the advantage of allowing estimation of levels of care, risk-factor management, and rates of outcomes, taking a broader perspective than is possible within a single database. Such analyses can lead to revaluation of medical practice methods, medication use, and the cost-effectiveness of specific interventions within each system.

Development of EMR-based registries by privately managed health systems has also provided an opportunity for large multi–health system aggregators such as IBM MarketScan Research, DARTNet, Optum, the Centers for Medicare & Medicaid Services, and others. Their databases contain information on billing claims for various services, pharmacy records, and laboratory data on large segments of the population. This information can be linked to other factors at the health plan level or to external information on geographic location and socioeconomic patterns. Thus, they can broaden the population included beyond that of individual health systems. However, these aggregating systems require substantial financial resources and can have other limitations. Although individual-level longitudinal analyses are possible with such systems, they can be complicated by the flow of individuals in and out of health plans, requiring careful distinction between cross-sections and cohorts. Aggregated health-system data also may lack routinely collected information on health behaviors and any information on the historically and geographically variable proportion of the U.S. population that is uninsured. Finally, as with databases within individual health systems, the completeness and reliability of aggregated data sets varies widely and poses significant problems of interpretation.
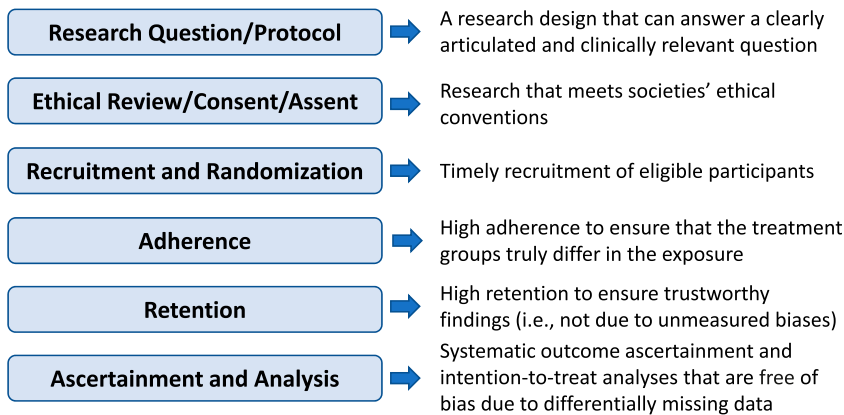
CLINICAL TRIAL DATABASES: OPTIMIZED COLLECTION AND ANALYSIS FOR SPECIFIC QUESTIONS

Complete and accurate quantitative data are required for success in all disciplines involved in scientific research. Clinical research can generally be classified as either observational or experimental. Observational research relies on data generated by people, clinics, institutions, health systems, or devices that are obtained, often passively, from sources such as an EMR system. Any variety of exposures, differences, or changes can be analyzed to identify relationships between the topic of investigation and various outcomes. Examples of topics for study include the uptake of a new drug, a change in health policy, an increase or decrease in access to health care providers, genetic characteristics, or increasing duration of disease or surveillance. Clinical assessments that can be related to such topics include weight or blood pressure, laboratory tests (e.g., A1C), health system utilization (e.g., emergency room visits), symptomatic events (e.g., hypoglycemia), and medical outcomes (e.g., myocardial infarction). All kinds of information collected during routine medical care might be used for observational research, and data are increasingly stored in easily accessible digital forms to facilitate their analysis.

Although observational research can identify relationships, whether any relationship is caused by the exposure or by something else linked to the exposure (i.e., a confounding variable) is more difficult to discern. Although sophisticated statistical techniques can account for potential confounding variables, they can only account for those that are both known to be possible confounders and available in the database. Because any relationship may reflect the effect of an unknown number of confounding factors, both measured and unmeasured, a causal effect suggested from observational analysis should be viewed as hypothesis-generating rather than definitive evidence. The only exception would be relationships that are extremely strong, such as the effect of smoking on the risk of lung cancer or the ability of insulin to prevent death in patients with type 1 diabetes.

Observational studies are no substitute for randomized controlled trials (RCTs) in establishing efficacy (72). The RCT is the gold standard for detecting modest but clinically important effects of a treatment or intervention. Indeed, a large number of RCTs conducted in the past 25 years (73) have provided crucial insights into the management of diabetes and have identified novel life-saving therapies. In an RCT, the administration of the exposure versus the comparator is randomly determined for two or more groups, and the effect of one versus alternate exposures (comparators) is then measured. The randomization process reduces confounding by constructing treatment groups that are, on average, expected to be similar except for the extent of the exposure being studied. Thus, any difference in outcomes is attributable to the exposure and not something else, with a level of confidence that depends on the rigor with which the study is designed and conducted. Many different exposures can be tested in RCTs, including drugs, devices, monitoring procedures (e.g., continuous glucose monitoring [CGM]), treatment algorithms, and administrative policies. Whereas the unit of randomization is typically an individual, groups or clusters of individuals can also be randomly assigned, with different clusters being randomly allocated to different exposures.

Although the methodological strengths of randomization are profound, randomization alone is not sufficient. Other requirements must be met (Fig. 3). First, a clearly formulated and ethical research question or hypothesis must be articulated as part of a carefully designed protocol. This should be reviewed by impartial experts to ensure that the question is important and that the research plan is ethical and feasible. Second, sufficient numbers of participants must be enrolled within a short enough period of time to ensure that the allocated groups are well matched and that the trial will be finished quickly enough to be relevant. Third, systems should be in place to ensure that people who are allocated to the exposure being tested actually adhere to or receive it. The lower the level of adherence, the smaller the difference between the allocated groups will be, so that a trial with low adherence may fail to detect very

Research Question/Protocol	⇒	A research design that can answer a clearly articulated and clinically relevant question
Ethical Review/Consent/Assent	⇒	Research that meets societies' ethical conventions
Recruitment and Randomization	⇒	Timely recruitment of eligible participants
Adherence	⇒	High adherence to ensure that the treatment groups truly differ in the exposure
Retention	⇒	High retention to ensure trustworthy findings (i.e., not due to unmeasured biases)
Ascertainment and Analysis	⇒	Systematic outcome ascertainment and intention-to-treat analyses that are free of bias due to differentially missing data

Figure 3—Six crucial components that ensure the robustness of a randomized controlled trial.

important effects of the exposure. Fourth, follow-up of study outcomes must be as close to 100% as possible to avoid the possibility that those who are not followed in each of the treatment groups may differ in important ways from those who are followed. Otherwise, these differences may confound the randomization and limit the confidence of conclusions about the effect of the exposure on the outcome. As illustrated in Table 2, most large-scale, randomized cardiovascular outcomes trials in diabetes have achieved follow-up rates for vital status approaching 100% (74–86). Fifth, systems need to be in place to ensure that outcomes of interest occurring during follow-up are reliably collected and analyzed. This is accomplished for very high percentages of participants in trials such as those shown in Table 2. Finally, the results should be analyzed according to the originally allocated exposure (i.e., through an intention-to-treat approach) regardless of adherence to the exposure.

RCTs have additional strengths. First, if two or more interventions work in very different ways, they can be tested at the same time in a large enough RCT. For example, in a 2-by-2 factorial design all participants in the study population are randomized to one intervention being tested or to its comparator and also to another intervention being tested or to its comparator. Such designs have been extremely successful and have seldom been undermined by unanticipated interactions between the therapies. Second, the database from an RCT can also be used for observational research. Indeed, any analyses done on the data that are not related to the comparison of the randomized treatment groups are essentially observational analyses through which associations can be explored. Third, substudies that rely upon collection of additional data (e.g., blood tests, images, or physiological measures) can also be built into trials to help determine the mechanism of action of the exposure (e.g., the effects of a new drug).

Because accuracy and completeness of data are centrally important to RCTs, procedures have been devised to ensure that the data collected are of the highest quality. Digital technology has been essential to this effort. Most RCTs have been done by specialized groups within specialized infrastructures erected to monitor and support the work of each trial (i.e., a coordinating center, a steering committee, multiple clinical sites, an event adjudication committee, and a data and safety monitoring board). Such infrastructures are complex and expensive, especially when there are many study participants who are

Table 2—Ascertainment of vital status in recent diabetes outcomes trials

Trial acronym (drug studied)	Patients randomized (n)	Median follow-up (years)	Vital status known (%)
ORIGIN (glargine & omega-3 fatty acid) (74)	12,537	6.2	99.0
SAVOR-TIMI 53 (saxagliptin) (75)	16,492	2.1	99.1
EXAMINE (alogliptin) (76)	5,308	1.5	99.5
TECOS (sitagliptin) (77)	14,671	3.0	97.5
EMPA-REG OUTCOME (empagliflozin) (78)	7,020	3.1	99.2
ELIXA (lixisenatide) (79)	6,068	2.1	99.0
LEADER (liraglutide) (80)	9,340	3.8	96.8
SUSTAIN-6 (semaglutide) (81)	3,297	2.1	99.6
CANVAS Program (canagliflozin) (82)	10,142	2.4	99.6
EXSCEL (exenatide) (83)	14,752	3.2	98.8
ACE (acarbose) (84)	6,522	5.0	94.4
HARMONY Outcomes (albiglutide) (85)	9,463	1.6	99.4
DECLARE-TIMI 58 (dapagliflozin) (86)	17,160	4.2	99.5

ACE, Acarbose Cardiovascular Evaluation; CANVAS, Canagliflozin Cardiovascular Assessment Study; DECLARE-TIMI 58, Dapagliflozin Effect on Cardiovascular Events; ELIXA, Evaluation of Lixisenatide in Acute Coronary Syndrome; EMPA-REG OUTCOME, BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients; EXAMINE, Examination of Cardiovascular Outcomes with Alogliptin versus Standard of Care; EXSCEL, Exenatide Study of Cardiovascular Event Lowering; HARMONY Outcomes, A Long Term, Randomized, Double-blind, Placebo-Controlled Study to Determine the Effect of Albiglutide, When Added to Standard Blood Glucose Lowering Therapies, on Major Cardiovascular Events in Patients With Type 2 Diabetes Mellitus; LEADER, Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results; ORIGIN, Outcome Reduction With Initial Glargine Intervention; SAVOR-TIMI 53, Saxagliptin Assessment of Vascular Outcomes Recorded in Patients with Diabetes Mellitus–Thrombolysis in Myocardial Infarction 53; TECOS, Trial Evaluating Cardiovascular Outcomes With Sitagliptin; SUSTAIN-6, Trial to Evaluate Cardiovascular and Other Long-term Outcomes With Semaglutide in Subjects With Type 2 Diabetes.

geographically dispersed. The personnel and procedures assembled for large RCTs are capable of collecting and rapidly processing massive amounts of meticulously collected data. For example, the ORIGIN (Outcome Reduction With Initial Glargine Intervention) trial gathered data for more than 6 years concerning more than 12,500 participants in 40 countries (74). Each trial's infrastructure is usually disbanded on completion of the trial.

Databases of well-conducted RCTs contain the highest-quality data, both to answer the primary question that prompted the study and to generate additional hypotheses from observational analyses. There are, however, some limitations to these data sets. They are usually focused on a very specific question that is addressed by measuring a specific primary outcome of interest and may not include observations or measurements that are relevant to some other questions. Furthermore, the population studied—which has been selected for being both able and willing to participate—may not be completely representative of the general population. Efforts are commonly made to analyze the data in such a way as to assess generalizability of the conclusions of the RCT, but questions often remain (87). The rapid growth of both EMRs and large public and private registries has the potential to address these problems. Specifically, such databases may facilitate enrollment of suitable study populations for randomized trials and also assist in tracking various measures of outcome (88).

OPPORTUNITIES FOR IMPROVING MEDICAL DATA MANAGEMENT

Although there are substantial variations between databases within each category, those derived from EMRs, those created from surveys and registries, and those created for RCTs or prospective

observational trials generally differ in their strengths and limitations. Some of their characteristics are summarized in Table 3.

Information in EMRs deals with large groups of individual patients, includes a comprehensive range of clinical material, is collected continuously, and is intended to be stored indefinitely. However, except in the case of certain countrywide health systems that provide care for all citizens, the population included in an EMR may not be fully representative of the general population. Interpretation of observations is generally limited by lack of consistency and accuracy in data entries. One important reason for this difficulty is that data are typically entered by many individual providers with little support or monitoring by administrative personnel. Further, the structure of EMRs differs widely among health systems and, thus, pooling or comparison of data from different systems is difficult.

Registries or surveys can include data that are consistently collected and often representative of a whole population. However, surveillance data sets may not be reliable in all cases (e.g., when collected by self-report), usually focus on one or a few categories of data, and may be updated only intermittently. When repeated cross-sectional information is collected, longitudinal observation of individuals may not be possible.

Databases created specifically for large clinical trials excel in the specificity, completeness, and accuracy of the data collected. However, they rarely include a fully representative sample of the patient population and may contain a relatively limited range of observations. They are also costly to build and maintain, and often are not maintained after completion of the trial.

An Evolving Approach: Distributed Data Networks

To some degree, the limitations of these several models for collecting and managing data can be overcome by development of comprehensive, distributed, multicenter data networks. Such "hub-and-spoke" models for linking individual databases differ from traditional multicenter studies or surveillance systems (in which all data are held centrally) in that individual data are maintained at their source, with analyses conducted peripherally using centrally coordinated common data models and analytic routines. Aggregate results standardized by the common data models and preestablished covariate adjustment can then be returned to the coordinating center for final analyses. Typically, such models have been used for comparative effectiveness studies of pharmacological options, bariatric surgery, and adverse events of drugs, but less often for surveillance of variation in risk, care, or outcomes of diabetes or for postmarketing drug surveillance programs (89–91). There are also opportunities for wider integration across care providers, including linkage of clinic EMRs with pharmacies to allow better monitoring of therapy adherence, and for automated acquisition of data from personal devices such as glucose monitors, insulin pumps and pens, exercise trackers, and health and fitness apps. One early example of regional EMR linkage for assessment of diabetes was the DARTS (Diabetes Audit and Research in Tayside, Scotland) study, which linked EMRs within a Scottish community to create a diabetes registry (92). More recently, groups of clinical investigators, such as the Blood Pressure Lowering Treatment Trialists' Collaboration and the Cholesterol Treatment Trialists' Collaboration, have formed for the purpose of aggregating individual data

Table 3—Attributes of current and potential future systems for managing medical data

Attributes	EMRs	Public surveys and registries	RCTs	Retrospectively aggregated data sets	Prospectively integrated data systems
Representative	++	+++	+	++	+++
Consistent	+	+++	+++	++	+++
Accurate	+	+	+++	+	+++
Comprehensive	+++	+	+	+	+++
Up-to-date	+++	+	+++	+	+++

+, ++, or +++ refer to the relative strength of each attribute, with +++ denoting the strongest. EMRs, electronic medical records; RCTs, randomized controlled trials.

from large trials (93,94). The possibility of expanding the range of data collection and analysis through such networks is obvious, but the quality of data can still be limited by inconsistencies and inaccuracies in clinical observations and data entry, and even aggregated data may not fully represent the general population.

A Potential Solution: Unified Data-Management Systems

In an ideal world, prospectively designed and unified data-management systems could support clinical, surveillance, and research activities all together in a way that circumvents many of the limitations of current systems while drawing on the strengths of each. An integrated system would, in theory, allow substantial savings in costs of design, development, operation, and maintenance, and these costs could be shared among multiple stakeholders.

Specifically, EMRs could be improved by incorporating the more stringent monitoring of data integrity, including automated validation of quantitative clinical and laboratory entries, which is typically used in trial-management systems. Population-wide surveillance of drug safety, rates of various adverse outcomes, regional differences in patterns of care, and other public health concerns could be based on improved and structured data collection during routine health care. Additionally, testing of new therapeutic agents, devices, or regimens could be embedded within existing health systems, using prospectively designed protocols for randomized or nonrandomized treatment choices and assessment of outcomes. Such an approach would facilitate enrollment of more representative and larger patient cohorts at lower cost. Patient follow-up and therapeutic adherence likely would be better if research studies were performed in a familiar usual-care setting.

Additional benefits of conducting RCTs or prospective observational studies using an EMR-based system would include the ability to follow patients passively long after the more structured initial part of the study. Long-term—even lifetime—individual follow-up could more fully capture the risks and benefits of the interventions evaluated and identify potential "legacy effects" persisting after completion of an active intervention. Beyond the opportunities related to individual trials and data cohorts, widespread implementation of unified data-management systems would facilitate routine analysis of pooled individual patient data, allowing greater representation of the whole population than is possible with current meta-analytic techniques. Access to such rich and long-term phenotypic information might also facilitate use of biobank and genetic data to identify new biomarkers and their relationships to disease outcomes and to existing and future therapies. Such unified big-data systems could provide a unique platform for testing, validating, and refining new analytic techniques such as artificial intelligence technologies, potentially leading to new diagnostic (95) and interventional tactics.

SUMMARY: MOVING TOWARD INTEGRATED SYSTEMS FOR MANAGING MEDICAL DATA

As noted above, networks that draw on varied sources are already accumulating experience with large, long-term aggregated data sets. For example, data collected over several decades in multiple population-based registries in both Norway and Sweden have been analyzed with the aim of improving regional health practices. These efforts have led to recent reports of marked and apparently continuing reduction of end-stage renal disease in type 1 diabetes during the period of surveillance (96,97). Also, a 5-year population-based intervention program in Hong Kong, based on structured EMR surveillance and decision-making by designated personnel trained in diabetes management, prospectively demonstrated large concurrent reductions of deaths, hospitalizations, and costs for patients with type 2 diabetes (98,99). Thus, movement toward integration of various kinds of health-related data is already underway.

As suggested by the examples above, the scale of the systems involved may be relevant to the success of integrated data management. Sweden, Norway, and Hong Kong all have populations in the 5- to 10-million range, and all have comprehensive publicly supervised health systems providing services to nearly all citizens. Fortunately, at present the computational power of electronic systems should not be a limiting factor in pursuing the goal of integrated data management for diabetes. However, the organizational and practical barriers to implementing integrated programs may be daunting over a larger geographic range and larger population than were demonstrated in these examples.

Several specific requirements appear to be necessary for implementing such systems in any setting. One is agreement on the definitions of key terms and goals in the management of diabetes, as for other medical conditions. Some progress has been made on this front for diabetes, as evidenced by consensus statements regarding glycemic measurements, glycemic targets, and hypoglycemia prompted by the recent development of CGM devices (100–102). Similarly, growing agreement on the properties and best uses of various glucose-lowering therapies is apparent in recent consensus statements by professional organizations (103,104). However, much remains to be done. Unified data systems will require national and international standardization of nomenclature and the incorporation of data dictionaries that can be used to encode diagnoses, procedures, drugs, and clinical outcomes. This may require building on established systems such as the Systematized Nomenclature of Medicine (SNOMED), International Classification of Diseases (ICD), and World Health Organization (WHO) classifications. Where required, mapping tools might be developed to allow translation of data sets among disparate coding systems. This is already being done by the National Library of Medicine, which maps the ICD-9 Clinical Modification, ICD-10 Clinical Modification, ICD-10 Procedure Coding System, and other classification systems to SNOMED, with the goal of establishing a universal taxonomy.

Additional difficulties are posed by proprietary concerns of competing health systems, hardware and software manufacturers, and data-management groups. Sharing of data and agreement on standardization of systems among businesses that are competing in the same markets may pose a significant barrier. In addition, security of protected health information must be ensured, and procedures to accomplish it must be agreed upon by various stakeholders.

However, there is precedent for resolving difficulties such as these. Standardization of electronic systems,

definitions, procedures, and regulations allowed for the development of international telephone service in the last century, and more recently the mechanics of the Internet and the World Wide Web. There seems no reason to believe that greater integration of data-management systems to facilitate diabetes care, surveillance, and research cannot be attained, given the potential for simultaneously improving medical outcomes and reducing overall costs.

In summary, integrated and improved management of big data has the potential to open a brave new world for diabetes care and research. Already we see successful proof-of-concept efforts, but further progress depends on overcoming logistical, administrative, and ethical obstacles to linking currently separate data-based activities.

References

1. Wiederhold G. Database technology in health care. J Med Syst 1981;5:175–196
2. Faltermayer EK. Better care at less cost without miracles. Medical College of Virginia Quarterly 1970;6:111–113
3. Andrews RD, Beauchamp C. A clinical database management system for improved integration of the Veterans Affairs Hospital Information System. J Med Syst 1989;13:309–320
4. Chantler C, Clarke T, Granger R. Information technology in the English National Health Service. JAMA 2006;296:2255–2258
5. National Center for Health Statistics. National Health and Nutrition Examination Survey history [Internet], 14 November 2011. Available from www.cdc.gov/nchs/nhanes/history.htm. Accessed 15 January 2019
6. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837–853
7. Nathan DM, Genuth S, Lachin J, et al.; Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993;329:977–986
8. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2016;(Suppl. 1):S48–S61
9. Nightingale F. *Notes on Hospitals*. 3rd Ed. London, England, Longman, Green, Longman, Roberts, and Green, 1863
10. Kucher N, Koo S, Quiroz R, et al. Electronic alerts to prevent venous thromboembolism among hospitalized patients. N Engl J Med 2005;352:969–977
11. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012;157:29–43
12. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The effect of electronic prescribing on medication errors and adverse drug events: a systematic review. J Am Med Inform Assoc 2008;15:585–600
13. Dorr D, Bonner LM, Cohen AN, et al. Informatics systems to promote improved care for chronic illness: a literature review. J Am Med Inform Assoc 2007;14:156–163
14. Cebul RD, Love TE, Jain AK, Hebert CJ. Electronic health records and quality of diabetes care. N Engl J Med 2011;365:825–833
15. Reed M, Huang J, Graetz I, et al. Outpatient electronic health records and the clinical care and outcomes of patients with diabetes mellitus. Ann Intern Med 2012;157:482–489
16. Reed M, Huang J, Brand R, et al. Implementation of an outpatient electronic health record and emergency department visits, hospitalizations, and office visits among patients with diabetes. JAMA 2013;310:1060–1065
17. McDonald CJ. Computerization can create safety hazards: a bar-coding near miss. Ann Intern Med 2006;144:510–516
18. Koppel R, Metlay JP, Cohen A, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005;293:1197–1203

19. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic health record usability issues and potential contribution to patient harm. JAMA 2018;319:1276–1278
20. Sittig DF, Wright A, Ash J, Singh H. New unintended adverse consequences of electronic health records. Yearb Med Inform 2016;Nov. 10:7–12
21. Han YY, Carcillo JA, Venkataraman ST, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics 2005;116:1506–1512
22. Margalit RS, Roter D, Dunevant MA, Larson S, Reis S. Electronic medical record use and physician-patient communication: an observational study of Israeli primary care encounters. Patient Educ Couns 2006;61:134–141
23. Street RL Jr, Liu L, Farber NJ, et al. Provider interaction with the electronic health record: the effects on patient-centered communication in medical encounters. Patient Educ Couns 2014;96:315–319
24. Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10:523–530
25. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful optimization of large-scale health information technology. J Am Med Inform Assoc 2017;24:182–187
26. Kaipio J, Lääveri T, Hyppönen H, et al. Usability problems do not heal by themselves: national survey on physicians' experiences with EHRs in Finland. Int J Med Inform 2017;97:266–281
27. Topaz M, Ronquillo C, Peltonen LM, et al. Nurse informaticians report low satisfaction and multi-level concerns with electronic health records: results from an international survey. AMIA Annu Symp Proc 2017;2016:2016–2025
28. Payne TH, Corley S, Cullen TA, et al. Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. J Am Med Inform Assoc 2015;22:1102–1110
29. Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011;44:1056–1067
30. Koppel R, Kreda D. Health care information technology vendors' "hold harmless" clause: implications for patients and clinicians. JAMA 2009;301:1276–1278
31. Goodman KW, Berner ES, Dente MA, et al.; AMIA Board of Directors. Challenges in ethics, safety, best practices, and oversight regarding HIT vendors, their customers, and patients: a report of an AMIA special task force. J Am Med Inform Assoc 2011;18:77–81
32. Ratwani RM, Hodgkins M, Bates DW. Improving electronic health record usability and safety requires transparency. JAMA 2018;320:2533–2534
33. Tahir D. Doctors barred from discussing safety glitches in US-funded software. Politico,11 September 2015 [article online]. Available from www.politico.com/story/2015/09/doctors-barred-from-discussing-safety-glitches-in-us-funded-software-213553. Accessed 17 January 2019
34. Coorevits P, Sundgren M, Klein GO, et al. Electronic health records: new opportunities for clinical research. J Intern Med 2013;274:547–560

35. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Use of electronic health record data in clinical investigations: guidance for industry. July 2018. Available from www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM501068.pdf. Accessed 17 January 2019

36. Brown JB, Nichols GA. Slow response to loss of glycemic control in type 2 diabetes mellitus. Am J Manag Care 2003;9:213–217

37. Brown JB, Nichols GA, Perry A. The burden of treatment failure in type 2 diabetes. Diabetes Care 2004;27:1535–1540

38. Whitmer RA, Karter AJ, Yaffe K, Quesenberry CP Jr, Selby JV. Hypoglycemic episodes and risk of dementia in older patients with type 2 diabetes mellitus. JAMA 2009;301:1565–1572

39. De Moor G, Sundgren M, Kalra D, et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015;53:162–173

40. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21:578–582

41. Visweswaran S, Becich MJ, D'Itri VS, et al. Accrual to Clinical Trials (ACT): a clinical and translational science award consortium network. JAMIA Open 2018;1:147–152

42. Mastellos N, Andreasson A, Huckvale K, et al. A cluster randomised controlled trial evaluating the effectiveness of eHealth-supported patient recruitment in primary care research: the TRANS-FoRm study protocol. Implement Sci 2015;10:15

43. Desai J, Geiss L, Mukhtar Q, et al. Public health surveillance of diabetes in the United States. J Public Health Manag Pract 2003; (Suppl.):S44–S51

44. Ali MK, Siegel KR, Laxy M, Gregg EW. Advancing measurement of diabetes at the population level. Curr Diab Rep 2018;18:108

45. Toh S, Platt R. Is size the next big thing in epidemiology? Epidemiology 2013;24:349–351

46. Centers for Disease Control and Prevention. Diabetes home: data and statistics [Internet]. 2019. Available from www.cdc.gov/diabetes/statistics/index.htm. Accessed 12 April 2019

47. Carstensen B, Kristensen JK, Marcussen MM, Borch-Johnsen K. The National Diabetes Register. Scand J Public Health 2011;39(Suppl):58–61

48. Birkner R. Plan and initial program of the Health Examination Survey. Vital Health Stat 1 1965;3:1–43

49. National Center for Health Statistics. Plan and operation of the Health and Nutrition Examination Survey: United States—1971-1973 [Internet]. Available from: https://www.cdc.gov/nchs/data/series/sr_01/sr01_010a.pdf. Accessed 12 April 2019

50. National Center for Health Statistics. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-94. Series 1: programs and collection procedures. Vital Health Stat 1 1994;32:1–407

51. National Center for Health Statistics. NHANES 1999–2000: Data, documentation, codebooks, SAS code [Internet]. Available from wwwn.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=1999. Accessed 18 January 2019

52. Centers for Disease Control and Prevention. Ambulatory health care data. Available from www.cdc.gov/nchs/ahcd/index.htm. Accessed 14 March 2019

53. Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey. Available from meps.ahrq.gov/mepsweb. Accessed 14 March 2019

54. Healthcare Cost and Utilization Project. Overview of the National (Nationwide) Inpatient Sample (NIS) [Internet], 2018. Available from https://www.hcup-us.ahrq.gov/nisoverview.jsp. Accessed 18 January 2019

55. Collins AJ, Foley RN, Gilbertson DT, Chen SC. United States Renal Data System public health surveillance of chronic kidney disease and end-stage renal disease. Kidney Int Suppl (2011) 2015; 5:2–7

56. Dabelea D, Mayer-Davis EJ, Saydah S, et al.; SEARCH for Diabetes in Youth Study. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. JAMA 2014;311:1778–1786

57. NHS Digital. National Diabetes Audit. Available from digital.nhs.uk/data-and-information/clinical-audits-and-registries/national-diabetes-audit. Accessed 6 February 2019

58. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017. Atlanta, GA, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, 2017

59. Zhang Y, Huang J, Wang P. A prediction model for the peripheral arterial disease using NHANES data. Medicine (Baltimore) 2016;95:e3454

60. Zhang X, Saaddine JB, Chou C-F, et al. Prevalence of diabetic retinopathy in the United States, 2005-2008. JAMA 2010;304:649–656

61. Murphy D, McCulloch CE, Lin F, et al.; Centers for Disease Control and Prevention Chronic Kidney Disease Surveillance Team. Trends in the prevalence of chronic kidney disease in the United States. Ann Intern Med 2016;165:473–481

62. Geiss LS, Wang J, Cheng YJ, et al. Prevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, United States, 1980-2012. JAMA 2014;312:1218–1226

63. Centers for Disease Control and Prevention. National Health Interview Survey: questionnaires, datasets, and related documentation [Internet]. Available from www.cdc.gov/nchs/nhis/nhis_questionnaires.htm2015. Accessed 20 April 2017

64. Gregg EW, Li Y, Wang J, et al. Changes in diabetes-related complications in the United States, 1990-2010. N Engl J Med 2014;370:1514–1523

65. Centers for Disease Control and Prevention (CDC). Estimated county-level prevalence of diabetes and obesity - United States, 2007. MMWR Morb Mortal Wkly Rep 2009;58:1259–1263

66. National Center for Health Statistics. NCHS data linked to NDI morbidity files [Internet]. Available from www.cdc.gov/nchs/data-linkage/mortality.htm. Accessed 18 January 2019

67. Engelgau MM, Geiss LS, Manninen DL, et al.; CDC Diabetes in Managed Care Work Group. Use of services by diabetes patients in managed care organizations. Development of a diabetes surveillance system. Diabetes Care 1998;21:2062–2068

68. Selby JV, Ray GT, Zhang D, Colby CJ. Excess costs of medical care for patients with diabetes in a managed care population. Diabetes Care 1997;20:1396–1402

69. Gudbjörnsdottir S, Cederholm J, Nilsson PM, Eliasson B; Steering Committee of the Swedish National Diabetes Register. The National Diabetes Register in Sweden: an implementation of the St. Vincent Declaration for Quality Improvement in Diabetes Care. Diabetes Care 2003;26:1270–1276

70. Niemi M, Winell K. Diabetes in Finland: Prevalence and Variation in Quality of Care. Available from http://www.diabetes.fi/files/1105/Diabetes_in_Finland._Prevalence_and_Variation_in_Quality_of_Care.pdf. Accessed 6 February 2019

71. Newton J, Garner S. Disease Registers in England: a Report Commissioned by the Department of Health Policy Research Programme in Support of the White Paper Entitled Saving Lives: Our Healthier Nation. Oxford, U.K., Institute of Health Sciences, University of Oxford, 2002

72. Gerstein HC, McMurray J, Holman RR. Real-world studies no substitute for RCTs in establishing efficacy. Lancet 2019;393:210–211

73. Cefalu WT, Kaul S, Gerstein HC, et al. Cardiovascular outcomes trials in type 2 diabetes: where do we go from here? Reflections from a Diabetes Care Editors' Expert Forum. Diabetes Care 2018;41:14–31

74. Gerstein HC, Bosch J, Dagenais GR, et al.; ORIGIN Trial Investigators. Basal insulin and cardiovascular and other outcomes in dysglycemia. N Engl J Med 2012;367:319–328

75. Scirica BM, Bhatt DL, Braunwald E, et al.; SAVOR-TIMI 53 Steering Committee and Investigators. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. N Engl J Med 2013;369:1317–1326

76. White WB, Cannon CP, Heller SR, et al.; EXAMINE Investigators. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. N Engl J Med 2013;369:1327–1335

77. Green JB, Bethel MA, Armstrong PW, et al.; TECOS Study Group. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. N Engl J Med 2015;373:232–242

78. Zinman B, Wanner C, Lachin JM, et al.; EMPA-REG OUTCOME Investigators. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med 2015;373:2117–2128

79. Pfeffer MA, Claggett B, Diaz R, et al.; ELIXA Investigators. Lixisenatide in patients with type 2 diabetes and acute coronary syndrome. N Engl J Med 2015;373:2247–2257

80. Marso SP, Daniels GH, Brown-Frandsen K, et al.; LEADER Steering Committee; LEADER Trial Investigators. Liraglutide and cardiovascular outcomes in type 2 diabetes. N Engl J Med 2016;375:311–322

81. Marso SP, Bain SC, Consoli A, et al.; SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. N Engl J Med 2016;375:1834–1844

82. Neal B, Perkovic V, Mahaffey KW, et al.; CANVAS Program Collaborative Group. Canagliflozin and cardiovascular and renal events in type 2 diabetes. N Engl J Med 2017;377:644–657

83. Holman RR, Bethel MA, Mentz RJ, et al.; EXSCEL Study Group. Effects of once-weekly exenatide on cardiovascular outcomes in

type 2 diabetes. N Engl J Med 2017;377:1228–1239

84. Holman RR, Coleman RL, Chan JCN, et al.; ACE Study Group. Effects of acarbose on cardiovascular and diabetes outcomes in patients with coronary heart disease and impaired glucose tolerance (ACE): a randomised, double-blind, placebo-controlled trial. Lancet Diabetes Endocrinol 2017;5:877–886

85. Hernandez AF, Green JB, Janmohamed S, et al.; Harmony Outcomes committees and investigators. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. Lancet 2018;392:1519–1529

86. Wiviott SD, Raz I, Bonaca MP, et al.; DECLARE–TIMI 58 Investigators. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. N Engl J Med 2019;380:347–357

87. Kosiborod M, Lam CSP, Kohsaka S, et al.; CVD-REAL Investigators and Study Group. Cardiovascular events associated with SGLT-2 inhibitors versus other glucose-lowering drugs: the CVD-REAL 2 study. J Am Coll Cardiol 2018;71:2628–2639

88. Bowman L, Mafham M, Stevens W, et al.; ASCEND Study Collaborative Group. ASCEND: A Study of Cardiovascular Events iN Diabetes: characteristics of a randomized trial of aspirin and of omega-3 fatty acid supplementation in 15,480 people with diabetes. Am Heart J 2018;198:135–144

89. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. Med Care 2010;48(Suppl.):S45–S51

90. Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. Ann Intern Med 2009;151:341–344

91. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. Health Aff (Millwood) 2014;33:1178–1186

92. Morris AD, Boyle DIR, MacAlpine R, et al.; DARTS/MEMO Collaboration. The Diabetes Audit and Research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. BMJ 1997;315:524–528

93. Ninomiya T, Perkovic V, Turnbull F, et al.; Blood Pressure Lowering Treatment Trialists' Collaboration. Blood pressure lowering and major cardiovascular events in people with and without chronic kidney disease: meta-analysis of randomised controlled trials. BMJ 2013;347:f5680

94. Kearney PM, Blackwell L, Collins R, et al.; Cholesterol Treatment Trialists' (CTT) Collaborators. Efficacy of cholesterol-lowering therapy in 18,686 people with diabetes in 14 randomised trials of statins: a meta-analysis. Lancet 2008;371:117–125

95. van der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem A, Nijpels G. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. Acta Ophthalmol 2018;96:63–68

96. Gagnum V, Stene LC, Leivestad T, Joner G, Skrivarhaug T. Long-term mortality and end-stage renal disease in a type 1 diabetes population diagnosed at age 15–29 years in Norway. Diabetes Care 2017;40:38–45

97. Toppe C, Möllsten A, Waernbaum I, et al.; Swedish Childhood Diabetes Study Group and the Swedish Renal Register. Decreasing cumulative incidence of end-stage renal disease in young patients with type 1 diabetes in Sweden: a 38-year prospective nationwide study. Diabetes Care 2019;42:27–31

98. Wan EYF, Fung CSC, Jiao FF, et al. Five-year effectiveness of the multidisciplinary Risk Assessment and Management Programme–Diabetes Mellitus (RAMP-DM) on diabetes-related complications and health service uses: a population-based and propensity-matched cohort study. Diabetes Care 2018;41:49–59

99. Jiao FF, Fung CSC, Wan EYF, et al. Five-year cost-effectiveness of the multidisciplinary Risk Assessment and Management Programme-Diabetes Mellitus (RAMP-DM). Diabetes Care 2018;41:250–257

100. Petrie JR, Peters AL, Bergenstal RM, Holl RW, Fleming GA, Heinemann L. Improving the clinical value and utility of CGM systems: issues and recommendations: a joint statement of the European Association for the Study of Diabetes and the American Diabetes Association Diabetes Technology Working Group. Diabetes Care 2017;40:1614–1621

101. Danne T, Nimri R, Battelino T, et al. International consensus on use of continuous glucose monitoring. Diabetes Care 2017;40:1631–1640

102. Agiostratidou G, Anhalt H, Ball D, et al. Standardizing clinically meaningful outcome measures beyond HbA$_{1c}$ for type 1 diabetes: a consensus report of the American Association of Clinical Endocrinologists, the American Association of Diabetes Educators, the American Diabetes Association, the Endocrine Society, JDRF International, The Leona M. and Harry B. Helmsley Charitable Trust, the Pediatric Endocrine Society, and the T1D Exchange. Diabetes Care 2017;40:1622–1630

103. Davies MJ, D'Alessio DA, Fradkin J, et al. Management of hyperglycemia in type 2 diabetes, 2018: a consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2018;41:2669–2701

104. Garber AJ, Abrahamson MJ, Barzilay JI, et al. Consensus statement by the American Association of Clinical Endocrinologists and American College of Endocrinology on the comprehensive type 2 diabetes algorithm—2018 executive summary. Endocr Pract 2018;24:91–120