Editorial

To the Challenges of Developing Large Diabetes Data Sets and Registries

Registries are known as data sets that contain health and demographic information. Large registries have the potential to inform best-care practices and to provide an opportunity for population surveillance and advocacy activities [1], which can help to reduce health inequalities. Registries are of interest to health-care providers, researchers and policymakers, and they are also relevant to people living with the condition. Although we recognize the importance of having access to such valuable resources, developing large data sets is not without challenges. In this issue of the journal, 3 groups of authors—Mathe et al [2], Sellers et al [3] and Ayub et al [4]—share their work in this quest to build accessible population representative data sets that can inform on diabetes care and decision-making.

One common challenge is the representativeness of the population included in the data set. Although diabetes cases using administrative databases are often defined using physician claims and hospital discharge information using the *International Classification of Diseases*, Mathe et al [2] propose to improve upon this definition by adding medication use and laboratory data. Their enhanced case definition resulted in an increased prevalence of diabetes, but, more importantly, the subsets of diabetes cases identified by each data source were complementary, and thus support that using a combination of data sources may improve case detection.

Another challenge is to identify different types of diabetes or subgroups of people with diabetes. For example, identifying those at high risk of diabetes may help improve health trajectories. Sellers et al [3] worked toward a more sensitive and specific definition of pregestational diabetes for use within administrative databases, given the increased risk for type 2 diabetes in the offspring. They have shown that the best administrative definition for pregestational diabetes exposure was any incident code for diabetes before pregnancy or within the first 25 weeks gestation. Their definition reveals that the prevalence of pregestational diabetes may have been underestimated using previous definitions that limited to a diagnosis before pregnancy or before 20 weeks gestation.

Although both Mathe et al and Sellers et al proposed to improve case definitions for application in administrative databases to generate data sets, Ayub et al [4] aim to build a provincial pediatric diabetes registry by enrolling patients in clinic, capturing harmonized patient-level (e.g. sociodemographic information, presentation at onset) and visit-level (e.g. glycated hemoglobin and weight) data, and to link, at some point, the collected information with other databases. Their challenges in developing a large data set were thus identified elsewhere—individual consent is required, and standardized procedures need to be established to ensure the accuracy and quality of the data collected (e.g. data abstraction and transfer from different paper and electronic medical records). However, their proposed strategy seemed to be welcomed by people living with diabetes. With a growing number of participants included, >85% of the approached patients consented to be part of this registry and, of these, 88% agreed to have their information linked with other registries (i.e. SWEET, an international pediatric registry) and with administrative databases. This shows that people living with diabetes also value the importance of registries.

The use of electronic health records and linkage to administrative databases provides a wealth of information and yields an evolving picture of health data. However, this does not capture the complete experience of people living with diabetes, which is multidimensional, including physical, psychological and social aspects. One main limitation is often the lack of patient-reported outcomes, such as psychological well-being and diabetes distress [5], which are key to understanding patients' concerns and experiences beyond clinical measurements. With careful planning, patient-reported outcomes can be included in registries, as proposed by Ayub et al [4]. In the process of building registries and using patient data, patients' values and preferences must be respected. We need to balance the need for direct patient engagement to achieve data acquisition, which may feel burdensome for some people, with the direct and indirect benefits patients will receive from having their experience and data shared.

Establishing registries—whether these include patients seen at the clinic or data from administrative databases—is associated with several challenges, which span from non-standardized electronic medical records and data billing errors to limited inclusion of patient-reported outcomes. By combining different approaches for recruitment or case definition and establishing harmonized minimal standard outcomes, we may succeed in building accurate real-world data sets with the potential to improve the care, health and longevity of people at risk of and living with diabetes.

Anne-Sophie Brazeau RD, PhD
Tricia Peters MD, PhD
McGill University, Montréal, Québec, Canada
E-mail address: anne-sophie.brazeau@mcgill.ca

Author Disclosures

References

1. Bak JCG, Semé EH, Kramer MHH, Nieuwdrop M, Verheugt CL. National diabetes registries: Do they make a difference? Acta Diabetol 2021;58:267–78.
2. Mathe N, Ryan A, Cook A, et al. Enhancing Diabetes Surveillance Across Alberta by Adding Laboratory and Pharmacy Data to the National Diabetes Surveillance System Methods. Can J Diabetes 2022;46:375–80.
3. Sellers EAC, Tan Q, Prior H, Chateau D, Wicklow BA, Fransoo R. Pregestational Diabetes Exposure In Utero: Validation of a Definition for Use in Administrative Data. Can J Diabetes 2022;46:388–91.
4. Ayub A, Ng C, Portales-Casamar E, Metzger D, Amed S. Toward Building a Provincial Diabetes Registry of Children and Youth Living With Diabetes in British Columbia, Canada. Can J Diabetes 2022;46:346–52.
5. Nano J, Carinci F, Okunade O, et al. A standard set of person-centred outcomes for diabetes mellitus: Results of an international and unified approach. Diabet Med 2020;37:2009–18.