Preventive Medicine 58 (2014) 17–21

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Risk distribution and its influence on the population targets for diabetes prevention☆

Laura C. Rosella [a,b,c,*], Michael Lebenbaum [a], Ye Li [a,c], Jun Wang [a], Douglas G. Manuel [a,b,d,e,f]

[a] Public Health Ontario, Toronto, Ontario, Canada
[b] Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada
[c] Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
[d] Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
[e] Department of Family Medicine and Epidemiology and Community Medicine, University of Ottawa, Canada
[f] Statistics Canada, Ottawa, Ontario, Canada

ARTICLE INFO

Article history:
Received 4 June 2012
Accepted 6 October 2013
Available online 23 October 2013

Keywords:
Diabetes mellitus, type 2
Risk assessment
Primary prevention

ABSTRACT

Objective. To quantify the influence of type 2 diabetes risk distribution on prevention benefit and apply a method to optimally identify population targets.

Methods. We used data from the 2011 Canadian Community Health Survey (N = 45,040) and the validated Diabetes Population Risk Tool to calculate 10-year diabetes risk. We calculated the Gini coefficient as a measure of risk dispersion. Intervention benefit was estimated using absolute risk reduction (ARR), number-needed-to-treat (NNT), and number of cases prevented.

Results. There is a wide variation of diabetes risk in Canada (Gini = 0.48) and with an inverse relation to risk (r = −0.99). Risk dispersion is lower among individuals meeting an empirically derived risk cut-off (Gini = 0.18). Targeting prevention based on a risk cut-off (10-year risk ≥ 16.5%) resulted in a greater number of cases prevented (340 thousand), higher ARR (7.7%) and lower NNT (13) compared to targeting individuals based on risk factor targets.

Conclusions. This study provides empirical evidence to demonstrate that risk variability is an important consideration for estimating the prevention benefit. Prioritizing target populations using an empirically derived cut-off based on a multivariate risk score will result in greater benefit and efficiency compared to risk factor targets.

Introduction

A key component to manage the burden of type 2 diabetes (T2DM) in the population is accurately identifying and characterizing baseline risk of developing T2DM in the population in order to appropriately plan and target prevention strategies. This includes articulating both the level of risk (likelihood of developing diabetes in the future) and the distribution of risk (what proportion of the population fall into a given risk category). The idea of risk dispersion was originally proposed by Rose, where he argued that variability of risk in the population can influence intervention effectiveness in terms of high-risk versus population-wide prevention (Rose, 1992). However, Rose's work focused on the conceptualization of risk conferred by a single risk factor (i.e. blood pressure). The use of predictive algorithms is an efficient approach to identifying risk cut-offs for targeted interventions that allows for the inclusion of multiple risk factors (McLaren et al., 2010). These approaches have recently been developed and validated for use at the population level (Manuel et al., 2012; Rosella et al., 2011).

While risk algorithms are increasingly being used in clinical and recently in population settings, further research is needed on how to best interpret and apply risk-cut-offs to inform intervention approaches. For example, it is not clear what magnitude of diabetes risk (e.g. 10-year risk ≥ 20%) would result in the greatest population benefit from a given diabetes prevention strategy. Most risk cut-offs identified from other algorithms appear arbitrary and are not designed to specifically maximize prevention outcomes. An important cut-off attribute that is currently missing from prevention strategies is maximizing strategy effic\acy, meaning the risk level used to identify target populations balances the number of individuals targeted with the potential benefit. In addition, few studies have directly examined how dispersion and concentration of diabetes risk in the population can influence the impact of a given strategy.

Abbreviations: ARR, absolute risk reduction; CCHS, Canadian Community Health Survey; DPoRT, Diabetes Population Risk Tool.

* Corresponding author at: Public Health Ontario, 480 University Avenue, Toronto, Ontario M5G 1V2, Canada.
E-mail address: laura.rosella@oahpp.ca (L.C. Rosella).

0091-7435/$ – see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ypmed.2013.10.007

The objectives of this study are to demonstrate how the dispersion of risk in the population, measured by the Gini coefficient, is correlated with the population risk of diabetes and to generate empiric risk cut-offs based on a validated risk score in order to maximize the population benefit as measured by absolute risk reduction in the population.

Methods

We first updated an existing validated risk prediction algorithm for incident diabetes, referred herein as DPoRT 2.0. DPoRT is a statistical model based on the Weibull survival distribution and is validated to calculate up to 10-year diabetes risk in any population-based data that contains self-reported risk factor information on age, height and weight, ethnicity, education, immigrant status, hypertension, self-reported heart disease, income, smoking and sex for those age 20 years and older and who are currently without diabetes. The original risk algorithm was based on a cohort of individuals 19,861 ≥ 20 years of age without diabetes followed between 1996 and 2005 and validated in two external cohorts in Ontario (N = 26,465) and Manitoba (N = 9899). Full details of development and validation can be found in a previous study (Rosella et al., 2011).

DPoRT 2.0 follows the same methodology with updated coefficients based on more recent data including individuals from the original 1996 Ontario cohort and the Ontario respondents of Cycle 1.1 (2001) and 2.1 (2003) of the Canadian Community Health Survey (CCHS) linked to the Ontario Diabetes Database (ODD) with follow-up until 2011 (Hux and Ivis, 2005) resulting in a total sample size of 69,606 individuals and 667,337 person-years of follow-up. DPORT 2.0 was externally validated in Ontario respondents to the 2005 CCHS linked to the ODD with follow-up until 2011. We examined two indices of model performance: discrimination and calibration. Model discrimination is the ability to correctly classify those with and without the disease based on predicted risk, i.e. correctly ranking those who will and will not develop diabetes. Discrimination is measured using a C statistic, which is analogous to the area under the receiver operating characteristic curve. This study uses a C statistic modified for survival data developed by Pencina and D'Agostino (2004). Calibration or accuracy is the extent of agreement between predicted and observed outcomes. It is measured using the Hosmer and Lemeshow statistic (H–L test), a χ^2 test, which measures observed and predicted values over deciles of predicted risk (D'Agostino et al., 2001; Hosmer and Lemenshow, 2000). In our study, it was calculated by comparing observed diabetes rates and DPoRT-predicted diabetes probabilities using a modified version of the H–L χ^2 statistic for time-to-event data (D'Agostino et al., 2001; Nam, 2000). To mark sufficient calibration, $\chi^2 = 20$ was used as a cut-off ($p < 0.01$). The CCHS is a nationally representative household survey of Canadians conducted by Statistics Canada which collects information on health status, determinants of health, and health care utilization. Households are selected though stratified, multilevel cluster sampling of private residences using provinces and/or local planning regions as the primary sampling unit. The surveys are conducted through telephone and in-person interviews and all responses are self-reported. The target population consists of persons aged 12 and over residing in private dwellings in all provinces and territories, except those living on Aboriginal reserves, on Canadian Forces Bases, or in some remote places. These surveys use a multistage stratified cluster design and provide cross-sectional data representative of 98% of the Canadian population over the age of 12 years. All surveys used for development, validation, and application of DPoRT attained at least a 75% overall response rate (Statistics Canada, 2002, 2003).

We applied the validated DPoRT 2.0 to Canadian adults (age ≥ 20), who are non-pregnant, free of diabetes and had valid information on risk factors in the 2011 CCHS Share file (N = 45,040). For every individual in the CCHS, we calculated 10-year diabetes risk and summarized this risk at the national level. We calculated confidence intervals taking into account both coefficient and complex survey variation generated using bootstrap techniques (Kovacevic et al., 2008).

The Gini coefficient applied to DPoRT-estimated risk was used as a measure of risk dispersion. The Gini coefficient is a measure of statistical dispersion (also known as variability) and can be simply defined as the average of all the absolute differences of pairs in a sample (Glasser, 1962). While typically used to describe income inequality, it is a general statistic of inequality that has been applied to a variety of other outcomes including other health indices (Asada, 2005). A Gini coefficient of zero expresses perfect equality where all values are the same for all individuals in a population (e.g. where everyone has exactly the same diabetes risk). A Gini coefficient of one expresses maximal inequality among values (e.g. where only one person has all the diabetes risk). We examine the relationship between level of risk in the population and dispersion of diabetes risk by ranking percentiles of the population and then calculating the Gini coefficient of the population included within percentile groups (e.g. 0.1 represents the top 10% of the population ordered by risk of diabetes). We plotted the relationship where the x axis represents sections taken from the population ranked from the highest diabetes risk to the lowest risk. As a greater proportion of the population is included, the average risk in that section of the population decreases given that lower risk groups are included. The y-axis represents the Gini coefficient for that section of the population. We then calculated the correlation coefficient of this relationship.

We examined how risk distribution measures would affect population intervention strategies by calculating the benefits of a hypothetical targeted intervention strategy using different approaches for identifying the target group that will receive the intervention. Specifically we quantified the impact of an intervention targeting the general population and high-risk groups based on single or dual risk factors (obesity and overweight among non-white ethnicities) or based on an empirically-derived risk cut-off estimated from DPoRT 2.0. We defined population benefit as the absolute risk reduction (ARR) in 10-year diabetes risk (absolute difference in diabetes risk before and after the intervention) and the corresponding number of diabetes cases prevented. The number of diabetes cases prevented was determined by summating the ARR multiplied by the survey weight for all targeted individuals. The Number Needed to Treat (NNT) is equal to one over the mean value of the ARR in the population. We based the effect of the diabetes prevention strategy on a plausible range seen from meta-analyses of intervention studies involving an intensive lifestyle intervention, typically a combination of diet and physical activity, which would have a larger effect on reducing 10-year diabetes risk (Gillies et al., 2008). For the intervention strategy we used a 10-year risk reduction of 30%; although, we examined a range of effect sizes (10–60%). We derived an optimal cut-point to identify a diabetes risk score that would identify individuals or groups that would benefit from intervention. The empirically-derived risk cut-off was based on a nonparametric discontinuity regression function that maximizes the difference in mean ARR between those who meet and do not meet the cut-point (Klotsche et al., 2009). This value is represented as solid black line in Fig. 2.

Results

The updated algorithm (DPoRT 2.0) demonstrates excellent accuracy (H–L $\chi^2 < 20$, $p < 0.01$?) and similar discrimination to the original DPoRT (C-statistic = 0.77) (Fig. 1) (Appendix A). Overall, based on the 2011 population, diabetes risk is 10% (9.6%, 10.4%) translating to over 2.25 million new diabetes cases expected in Canada between 2011 and 2020. The 10-year baseline risk for diabetes in the overall population and by important subgroups is reported in Table 1. Ten-year diabetes risk varies by age, Body Mass Index (BMI), sex, ethnicity, and quartile of risk. The absolute numbers of expected new cases reflect variation in risk across the population, in addition to distribution of sub-groups within the Canadian population.

Risk is variable in the Canadian population (Gini = 0.48); however, within subgroups there is a range of risk dispersions from as low as 0.11 to as high as 0.52 (Table 1). Diabetes risk is less variable within older ages, among those that are obese, and within quartiles of risk. High variability in 10-year diabetes risk is noted within certain ethnic groups and among those under 45.

The degree of variability in diabetes risk is related to the magnitude of diabetes risk such that the higher the diabetes risk score, the lower the dispersion among the population that falls below that risk cut-off ($r = -0.99$, Fig. 2). The empirically derived cut-off was determined to be a risk of 16.5% (Fig. 3). Table 2 demonstrates the benefit in targeting individual or dual risk factors compared to targeting based on an empirically derived risk cut-off. Risk dispersion is lower when using the empirically derived risk cut-off based on DPoRT compared to a single factor target, although they represent similar proportions of the population (20% vs. 17%). Furthermore, targeting the population that