original article

Estimating the prevalence of select non-communicable diseases in Saudi Arabia using a population-based sample: econometric analysis with natural language processing

Suliman Alghnam,[a] Mohammad Bosaeed,[b,c] Abdulrahman Aljouie,[d] Saeed Mastour Alshahrani,[e] Omar Alshenqeety,[f] Rifat Atun,[g] Saleh Alqahtani[h]

From the [a]Population Health Section, King Abdullah International Medical Research Center, Riyadh Saudi Arabia; ; [b]Department of Medicine, King Abdulaziz Medical City, Riyadh, Saudi Arabia; [c]College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia; [d]Department of Data Management, King Abdullah International Medical Research Center, Riyadh Saudi Arabia; [e]Department of Public Health, King Khalid University, Asir, Saudi Arabia; [f]Ministry of Health, Riyadh, Saudi Arabia; [g]Department of International Health, Harvard TH Chan School of Public Health, Boston, Massachusetts, United States; [h]Department of Hepatology, Johns Hopkins, Baltimore, Maryland, United States

Correspondence: Dr. Suliman Alghnam · Population Health Section, King Abdullah International Medical Research Center, Riyadh Saudi Arabia · alghnam.s@gmail.com · ORCID: https://orcid.org/0009-0000-9594-8408

Citation: Alghnam S, Bosaeed M, Aljouie A, Alshahrani SM, Alshenqeety O, Atun R, et al. Estimating the prevalence of select non-communicable diseases in Saudi Arabia using a population-based sample: econometric analysis with natural language processing. Ann Saudi Med 2024; 44(5): 329-338. DOI: 10.5144/0256-4947.2024.329

Received: July 1, 2024

Accepted: August 31, 2024

Published: October 3, 2024

Funding: None.

BACKGROUND: Non-communicable diseases (NCDs) are a major public health challenge globally, including in Saudi Arabia. However, measuring the true extent of NCD prevalence has been hampered by a paucity of nationally representative epidemiological studies.

OBJECTIVES: Assess the prevalence of selected NCDs, using population-based electronic health records and applying novel analytical methods to identify cases of NCDs.

DESIGN: Retrospective

SETTINGS: A large healthcare network in Saudi Arabia.

PATIENTS AND METHODS: We included all beneficiaries aged 16 years or older (n=650 835[a]) and used the International Classification of Disease (ICD-10) codes, laboratory results, and associated medications to identify individuals with diabetes, obesity, hypertension, dyslipidemia, mental disorders, and injuries. For diabetes and hypertension, we used natural language processing (NLP) on clinical notes in the electronic health records. The prevalence of multimorbidity across age groups was also tabulated, and logistic regression was used to examine its association with glycemic control.

MAIN OUTCOME MEASURES: The primary outcomes measured were the prevalence of diabetes, hypertension, and multimorbidity, and their association with glycemic control.

SAMPLE Size: 650 835 individuals aged 16 years or older.

RESULTS: The study population was relatively young, with 41.2% aged between 26 and 45 years, and around two-thirds were married. The prevalence of diabetes and hypertension was 18.5% (95% CI: 18.5-18.7) and 13.0% (95% CI: 12.9-13.1), respectively. Approximately 26.7% (95% CI: 26.7-26.8) of the population had multimorbidity, with levels increasing to 62.9% for those aged 65 or older. Multimorbidity was associated with a four-fold increase in the likelihood of poor glycemic control. NLP analysis suggested that the prevalence of diabetes or hypertension may be underestimated by no more than 1.5%.

CONCLUSIONS: The study suggests a higher prevalence of NCDs than earlier national estimates. Electronic health records with regular analysis provide an opportunity to estimate changes in the prevalence of NCDs in Saudi Arabia. Health policies and interventions are needed to address the high levels of multimorbidity, which adversely impact glycemic control.

LIMITATIONS: Retrospective design and reliance on electronic health records, which may not capture all cases of NCDs.

CONFLICT OF INTEREST: None.

The social and economic life in Saudi Arabia has changed substantially in the past seven decades.[1] Life expectancy increased from 55 years in 1972 to 75 in 2020, with improvements in other population health indicators. To further improve population health, the country launched the Saudi Vision 2030 in 2016, with one of the goals to increase life expectancy to 80 years by 2030.[2] Non-communicable diseases (NCDs) and injuries are the leading causes of death and disability in the country. Therefore, one of the major components of the health system transformation is to move from a curative model toward a value-based healthcare model that focuses on improving population health to enable better use of the investments in the health system. A primary objective of the transformation is to reduce the high burden of NCDs– with a 10% decrease in diabetes and a 3% reduction in obesity prevalence between 2016 and 2030.[3]

One of the challenges to reducing the NCD burden is the paucity of comprehensive, reliable, timely, and valid data on the prevalence of NCDs and health outcomes. Currently, most population health indicators are based on data from international estimates, such as the Global Burden of Disease (GBD) study.[4] A comprehensive assessment of the current NCD burden using local data is warranted in order to inform policies, plans, and investments to improve population health.

The population of Saudi Arabia is expected to reach 40 million by 2030, with an increase in the proportion of older age groups due to improvements in life expectancy.[2] With an aging population and without major public health interventions, NCDs will likely further increase, producing additional demand on the health system. Estimates by GBD suggest that four of the five leading causes of Disability Adjusted Life Years (DALYs) are high body mass index (BMI), high blood pressure, elevated fasting glucose, and high levels of low-density lipoprotein (LDL) cholesterol.[5] However, comprehensive real-world data (RWD) on these conditions in the country are too scarce to estimate the actual burden.[6]

Several conditions contribute to NCDs. Obesity is a major risk factor for cardiovascular diseases and other NCDs.[7] A study by Alhabib et al on the prevalence of cardiovascular risk factors in Saudi Arabia among those aged 35-70 years old found that one-half of this age group was obese, and 35.5% were overweight.8 Diabetes is another significant burden on the health system, with an estimated annual cost of about 0.87 billion US dollars, and is anticipated to increase in the next decade.[9] Therefore, the prevalence of NCDs must be assessed regularly to inform better planning and the development of suitable implementation prevention strategies.[10]

Estimates of the prevalence of NCDs in Saudi Arabia using local data vary, as these estimates are based on small samples, outdated reports, or different age cut-offs.[11] For example, a study in 2013 estimated that the prevalence of diabetes was 13.4%. In contrast, a national study in 2019 indicated that self-reported diabetes was 8.0%, despite an increase in many risk factors for diabetes (such as obesity) nationwide.[12,13] Hence, there is a need for 'real-time' or 'near real-time' monitoring of the prevalence of NCDs to inform decision-makers during health system transformation to guide policies, interventions, and resource allocation.

Electronic health records (EHR) can help provide timely data on disease prevalence, multimorbidity levels, healthcare utilization patterns, and health outcomes, as these data are captured regularly for individuals who use healthcare services. However, without screening programs, EHRs fail to capture data on individuals who may have an NCD but may not be aware of their condition. Cross-country studies in low-income and middle-income countries that have used nationally representative data with biomarkers indicate that the proportion of those with a major NCD, such as diabetes or hypertension who are not diagnosed may exceed 50%.[14] Nevertheless, EHR provides an accessible tool to all individuals receiving any healthcare service, including primary and prevention services, and may help monitor disease with minimal costs. Some countries have explored the use of EHR to estimate NCDs' prevalence and improve health outcomes.[15]

This study uses data of the beneficiaries from the EHR in a large health system, one of the major subsystems in the country with a comprehensive EHR system covering 3% of Saudi Arabia's population, to estimate the prevalence of NCDs and injuries.

PATIENTS AND METHODS

This is a retrospective study that utilizes data of the beneficiaries with records between 2016 and 2021. The study included all tertiary hospitals and outpatient clinics in a large healthcare network. The study included all fully eligible beneficiaries aged 16 years and older and for whom comprehensive data exist. Full eligibility was defined as full access to all medical and preventative services.

The study examined three categories of parameters. The first category was related to BMI (as a measure of obesity), diabetes, hypertension, dyslipidemia, injuries, and mental disorders. BMI is captured automatically in the system using weight (in kg) divided by height (in meters squared). Because the EHR includes all the weight and height measurements for every visit to a healthcare facility taken by registered nurses, the latest measurement was used. Subjects were categorized using the Center for Disease Prevention and Control as underweight (BMI <18.5), normal (BMI=18.5-24.9), overweight (BMI=25-29.9), or obese (≥30). Diabetic patients were defined as those diagnosed with diabetes (ICD-10 Codes=E10-E14; excluding gestational), taking hypoglycemic medication, having a hemoglobin A1c level ≥6.5, having fasting blood glucose >125, or having a random test of 200 or higher.[16]

Individuals with hypertension were either diagnosed as hypertensive (ICD-10 Codes=I10-I15) or taking antihypertensive medication. Dyslipidemia patients were defined as those with a diagnosis (ICD-10 Code=E78), taking medication, or having abnormal lab lipid levels. We defined abnormal lab as either: high total cholesterol (>5.19 mmol/L), high levels of LDL cholesterol (>2.61 mmol/L), low levels of high-density lipoprotein (HDL) cholesterol (males <1.04 or females <1.54 mmol/L), or triglyceride (>1.71 mmol/L). Injured patients were those seen in the emergency department, outpatient clinic, or admitted as inpatients following acute injuries (ICD-10 Codes=S00-T14). Patients with mental disorders were diagnosed based on the ICD-10 codes F0-F99.[17]

The second category of measurement is related to multimorbidity, defined as two or more conditions.[18] Health conditions included diabetes, hypertension, dyslipidemia, obesity, and mental disorders.

The third category of measurements is related to glycemic control among patients diagnosed with diabetes. Glycemic control was accessed using glycated hemoglobin levels and measured using hemoglobin A1c (HbA1c). We defined diabetes control as having a hemoglobin A1C level of <7 after the initial diagnosis.[19] Also, we defined persistent uncontrolled diabetes as more than five abnormal HbA1c (>7%) over the study period (range 1-5 years).

Calibration subsample

The EHR's clinical notes may specify disease diagnoses in unstructured formats. For example, a patient is diabetic and hypertensive and receives treatment in other facilities or private hospitals. Suppose this individual needed emergency care for an injury, and the physician documented that the individual has diabetes and hypertension in the clinical notes. In that case, these episodes of care provided in other networks may not be captured in the data system with the correct details of ICD codes, medications, or laboratory results. To mitigate this bias, we utilized a rule-based (with regular expression) natural language processing (NLP) system using Python scripting to validate the accurate detection of two conditions: i) diabetes and ii) hypertension.[20] To verify the true negatives for diabetes and hypertension in patients' data extracted from the EMR system, we randomly selected two subsamples from the population (~10 000 patients each)— one from those identified as non-diabetics and another from those who were identified as non-hypertensive. We called these calibration samples since we intended to capture the magnitude of the false negatives in the EHR system for those conditions. Calibration methods have been used previously to improve estimates in health research.[21] The data was retrieved from the system on February 21, 2022, and was deidentified prior to the analysis. The study was reviewed and approved by the respective Institutional Review Board.

Statistical analysis

Statistical analyses were conducted using STATA software (version 15 SE). Descriptive statistics and frequency tables were used to describe the demographics and prevalence of the disease. Age and gender-specific prevalence were calculated for the stated conditions and the distribution of diseases across the institution branches by country's geographic region.

We also examined and tabulated those with multimorbidity to understand its distribution in the study population. Conditions included diabetes, hypertension, dyslipidemia, obesity, and mental disorders. The association between multimorbidity and diabetes control was assessed using a Chi-square test. The frequency of uncontrolled HbA1c after the diagnosis date was tabulated and categorized into controlled (all HbA1c measurements following diagnosis were ≤7%), 1-2 high measurements, 3-5 high measurements, or >5 uncontrolled measurements (HbA1c level>7%). We also explored the association with persistent uncontrolled diabetes. This association was further modeled using a multivariable logistic regression model. The independent variables included age, gender, and whether the individual suffered from multimorbidity. Additional logistic regression models were constructed using different dyads of the chronic conditions studied.

For the calibration subsamples, we used Python scripting language for all analyses. We grouped patients' progress notes (encounters) into one aggregated text for each patient. Next, we evaluated every line (we treated each newline character as a sentence separator). A patient is considered non-diabetic if there is no mention of 'diabet' or 'DM' or "T1D" or "T2D" keywords or if the subsequence (within the same line) has one of the negation or family history contexts (**Appendix A**). Likewise, a patient is classified as non-hypertensive when the subsequent words does not contain 'hyperten' or 'htn' or "HTN" or these terms are negated.

RESULTS

The analysis included 650 835 beneficiaries (52.3% females and 47.6% males). The majority of the population sample was relatively young (26-45 years old, 41.1%), Saudi nationals (99.4%), married (58.8%), and residents of the central region (58.5%, **Table 1**). 67.4% of beneficiaries were overweight (28.9%; 95% CI=28.7-29.0) or obese (38.5%; 95% CI=38.3-38.6, **Table 2**). The prevalence of obesity was higher among those 46-64 than in other age groups (57.9%; 95% CI=57.7-58.3, **Figure 1**).

The results show that 18.5% (95% CI=18.5-18.7) of the study population had diabetes. Most patients (78.7%; 95% CI=78.5-78.9) were identified through the ICD codes. The remaining 21.3% of diabetic patients were identified via medication lists and lab records. They had not been diagnosed formally through the ICD code but had been either using diabetic medications or had elevated levels of HbA1c (≥6.5%), fasting blood glucose (>125 mg/dl), or random blood glucose (>200 mg/dl). The prevalence of diabetes was highest among those aged 65 years or more (53.2%; 95% CI=52.8-53.5, **Figure 2A**).

In the analysis, 13.0% (95% CI=12.9-13.1) of the study population had hypertension. Most hypertensive patients were identified using ICD codes (93%). The remaining 7% (95% CI=6.8-7.1) of hypertensive patients were identified after retrieving patients' medication lists who had not been diagnosed formally through the ICD code but had been using blood pressure-lowering medications (**Table 2**). As with diabetes, hypertension was high (48.7%, 95 CI=48.3-49.0) among those aged 65 years or older (**Figure 2B**).

As for dyslipidemia, we identified that 20.3% (95% CI=20.2-20.4) of the population had abnormal lipid profiles. Most were identified through the ICD code diagnosis documented in the HER system (83.5%), while the remaining 16.5% were identified by the patients' laboratory records and medication lists (**Table 2**). The prevalence of dyslipidemia was high among the 46-64 age group (41.0%; 95% CI=40.7-41.2; **Figure 2C**).

We identified that 80 995 injuries were treated during the study period within the healthcare network. In contrast to the distributions of diabetes, hypertension, and dyslipidemia across the age groups, injuries were more common among the younger population group (16-25 years, 15.9%, 95% CI=15.7-16.1: **Figure 2D**).

As for mental disorders, approximately 4.5% (95% CI=4.5-4.6) of the study population had a mental condition

Table 1. General characteristics of individuals participating in the health system (N=650 835).

Variable	
Age Group	
16-25	149 619 (22.99)
26-45	268 006 (41.18)
46-64	152 637 (23.45)
≥65	80 573 (12.38)
Gender	
Female	340 891 (52.38)
Male	309 944 (47.62)
Nationality	
Saudi	647 069 (99.42)
Non-Saudi	3766 (0.58)
Marital status	
Single	239 535 (36.8)
Married	383 074 (58.86)
Divorced/Separated/ Widowed	21 666 (3.33)
Unknown	6560 (1.01)
Region	
Central	381 159 (58.56)
Eastern	79 115 (12.16)
Western	154 417 (23.73)
Others	36 144 (5.55)

Data are number (percentage).

recorded (**Table 2**). Mainly, those aged 26-45 years had the highest prevalence of mental disorders (6.0%; 95% CI=5.8-6.1) than other age groups (**Figure 2E**).

Figure 3 illustrates the prevalence of multimorbidity in the study population. We found that 26.7% (95% CI=26.7-26.8) of the study population had multimorbidity with two or more chronic conditions, while 15.0% (95% CI=15.2-15.4) had three conditions or more. The prevalence of having three comorbidities or more was highest (48.8; 95% CI=48.5-49.2) among individuals 65 or older, representing around one-half of that age group.

Our results showed that less than one-third of the diabetic population achieved diabetes control (28.2%; 95% CI=28.0-28.4). Moreover, our results indicated a difference in diabetes control across age groups, with those aged 65 years or older having worse control than younger patients (*P* value <.01, **Figure 4**). This difference in control was also present when comparing individuals with just diabetes and with multimorbidity.

Those with multimorbidity were more likely to have persistent uncontrolled diabetes (>5 abnormal tests) over the study period than those with two or fewer comorbidities (**Table 4**). While over half of diabetic patients without comorbidities achieved glycemic control, 35.2% (95% CI=34.8%-35.5) of those with three or more chronic diseases had persistent uncontrolled diabetes. Regression analyses revealed a similar pattern with those with three or more chronic conditions and higher odds (OR=5.0; 95CI=4.7-5.3) of having persistent uncontrolled diabetes than diabetic patients without other comorbidities (**Table 3**). The dyad of diabetes and dyslipidemia had the highest level of president uncontrolled diabetes, with those who have dyslipidemia almost four times more likely to have persistent uncontrolled diabetes than those without comorbidities (OR=3.9; 95% CI=3.8-4.1).

Results from the calibration subsamples indicated that out of the 9968 (99.6%) out of the 10004 non-diabetic patients were true negatives and 36 (0.36%)

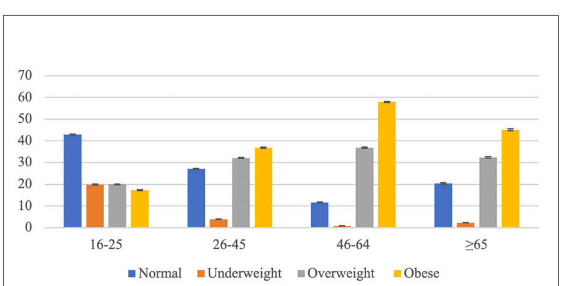

Figure 1. Body mass index class of National Guard Health Affairs beneficiaries across age groups (years).

Table 2. Diseases prevalence among beneficiaries of the health system (N=650 835).

Variable	
Diabetes	
No	529 944 [81.4 (81.3 81.5)]
Yes	120 891 [18.6 (18.5-18.7)]
Diabetes by the method of diagnosis	
ICD Code	95 134 [78.7 (78.5-78.9)]
No ICD Code but only high HbA1c[a]	2755 [2.3 (2.2-2.4)]
No ICD Code but only medication use	2935 [2.4 (2.3-2.5)]
No ICD Code but only high FBG[b]	9838 [8.1 (8.0-8.3)]
No ICD Code but only high RBG[c]	10 229 [8.5 (8.3-8.6)]
Hypertension	
No	566 210 [87.0 (86.9-87.1)]
Yes	84 625 [13.0 (12.9-13.1)]
Hypertension by the method of diagnosis	
ICD Code	78 723 [93.0 (92.9-93.2)]
No ICD Code but only medication use	5 902 [7.0 (6.8-7.1)]
Dyslipidaemia	
No	518 828 [79.7 (79.6-79.8)]
Yes	132 007 [20.3 (20.2-20.4)]
Dyslipidaemia by the method of diagnosis	
ICD Code	110 256 [83.5 (83.3-83.7)]
No ICD Code-only medication use	13 309 [10.1 (9.9-10.2)]
No ICD Code but only high lab[d]	8442 [6.4 (6.3-6.5)]
Injury[e,f,g]	
No	569 840 [87.6 (87.5-87.6)]
Yes	80 995 (12.4 [12.4-12.5)]
Mental disorders	
No	621 109 [95.4 (95.4-95.5)]
Yes	29 726 [4.6 (4.5-4.6)]
Body mass index	
Normal	134 765 [26.2 (26.1-26.3)]
Underweight	33 548 [6.5 (6.4-6.6)]
Overweight	148 563 [28.9 (28.7-29.0)]
Obese	198 070 [38.5 (38.3-38.6)]
Not reported/Not measured[h]	135 889

Data are number (percentage, 95% CI). [a]Glycated haemoglobin > 6.4%; [b]Fasting Blood Glucose > 125 mg/dL; [c]Random Blood Glucose >199 mg/dL; [d]Cholesterol; [e]Emergency Department admission; [f]In-Patient admission; [g]Out-Patient admission; [h]No weight data on the patients' records.

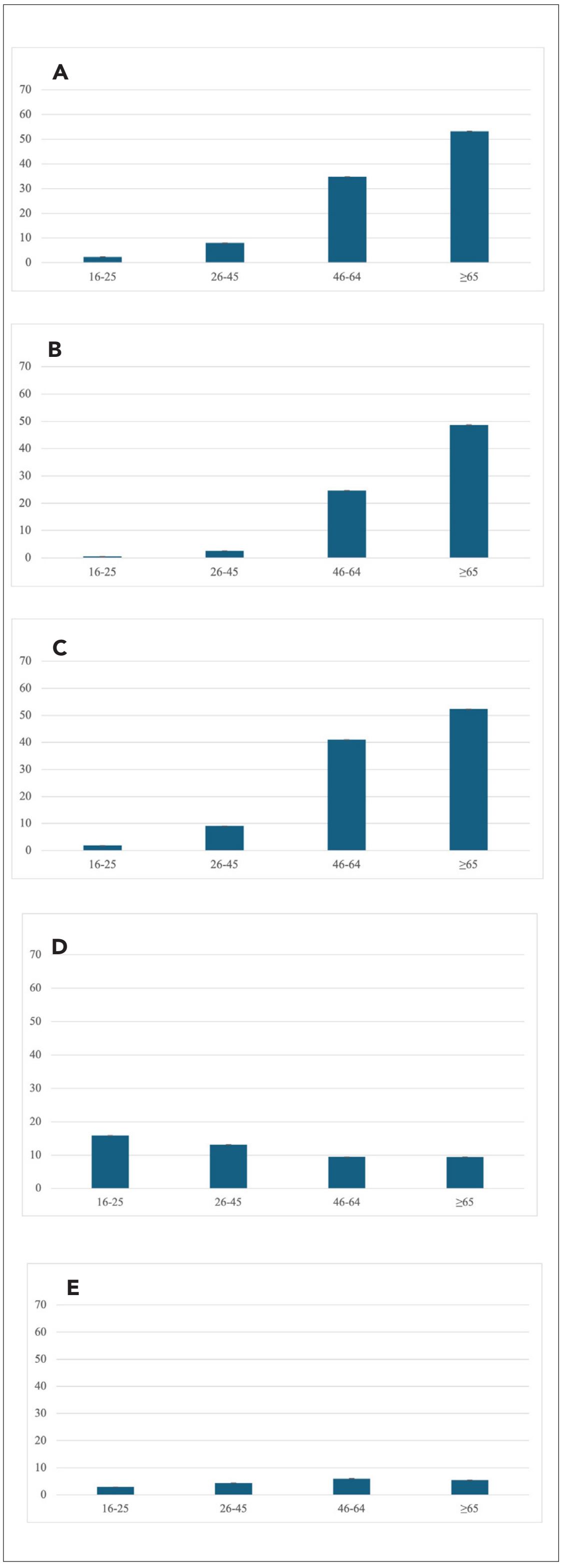

Figure 2. The prevalence of diseases in NGHA population by age groups (%): A) diabetes; B) hypertension; C) dyslipidemia; D) injury; E) mental disorders.

were false negatives based on the NLP approach. For hypertension, out of the randomly selected 10000 patients with extracted records without hypertension, the NLP system identified 9861(98.6%) as true negatives and 139 (1.4%%) patients as false negatives. All those identified as false negatives were further verified manually by reviewing the medical charts, and the findings were consistent with the analysis generated using NLP.

DISCUSSION

This analysis is the first approximation of a population-based sample in Saudi Arabia to quantify the prevalence of leading NCDs. Our results suggest that analysis using EHR-based data to monitor disease prevalence is feasible, and the findings could be used as part of health needs assessment to guide policies and interventions for population health management.

Our study found a high prevalence of four major NCDs and injuries among the study population. These levels are higher than earlier published estimates of the self-reported prevalence of diabetes to be around 8.0%.[13] Our finding of persistent uncontrolled diabetes is consistent with previously published local studies suggesting that three-fourths of patients with diabetes are not achieving control.[22] Addressing NCDs is a priority for Saudi Vision 2030, but a call for action to address the determinants of these conditions is desperately needed.

An important finding of the study is that more than 15% of the study population has three or more chronic conditions. Although this level is lower than the levels observed in high-income countries, where estimates suggest that around 31.7% of adults aged 20-64 had five or more chronic conditions, high-income countries are further advanced in epidemiological transition that brings chronic conditions than Saudi Arabia, which is in earlier stages of epidemiological transition.[23] However, around 50% of those aged 65 years or more had three or more chronic conditions. This level is higher than in upper-middle income countries, such as Indonesia, which are at a similar stage of epidemiological transition to Saudi Arabia, where only 22.0% of those 50 years or more had multimorbidity.[18] Patients with multimorbidity are of particular concern because they are likely to be the highest utilizer of healthcare resources and tend to have worse health outcomes and lower quality of life when compared with individuals with no multimorbidity.[23] Consequently, multimorbidity can affect individual prosperity and the nation's economy aside from healthcare utilization because those patients are more likely to be absent from work and incur high

Table 3. Multiple logistic regression models of the association between health comorbidities and poor glycemic control.

Variable	OR (95% CI)
Model 1: Age, gender and multimorbidity (n=120 891)	
Multimorbidity	
One	**Reference**
Two	**1.4 (1.3-1.5)**
>Two	**5.0 (4.7-5.3)**
Model 2 (Dyad #1): Age, gender and hypertension (n=33 030)	
Hypertension	
No	**Reference**
Yes	**1.5 (1.4-1.6)**
Model 3 (Dyad #2): Age, gender and dyslipidaemia (n=49 582)	
Dyslipidaemia	
No	**Reference**
Yes	**3.9 (3.8-4.1)**
Model 4 (Dyad #3): Age, gender and injury (n=29 957)	
Injury	
No	**Reference**
Yes	0.9 (0.8-1.0)
Model 5 (Dyad #4): Age, gender and mental condition (27 752)	
Mental condition	
No	**Reference**
Yes	1.0 (0.9-1.1)

BOLD=Significant at *P*<.01.

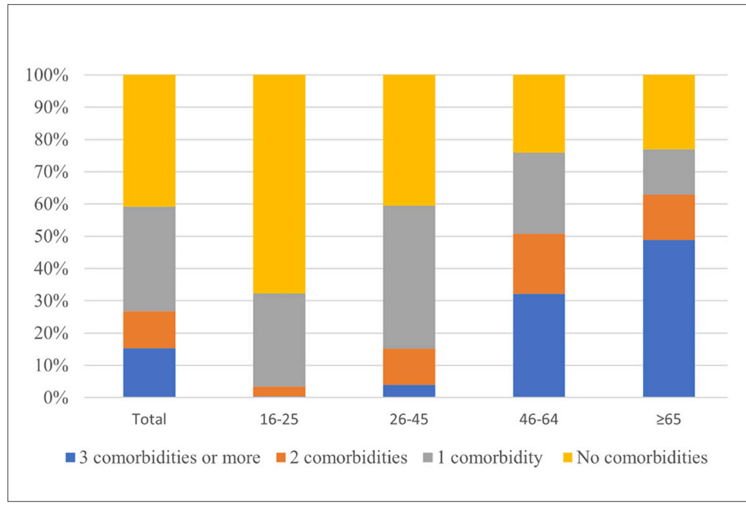

Figure 3. Prevalence of multimorbidity in the study population.

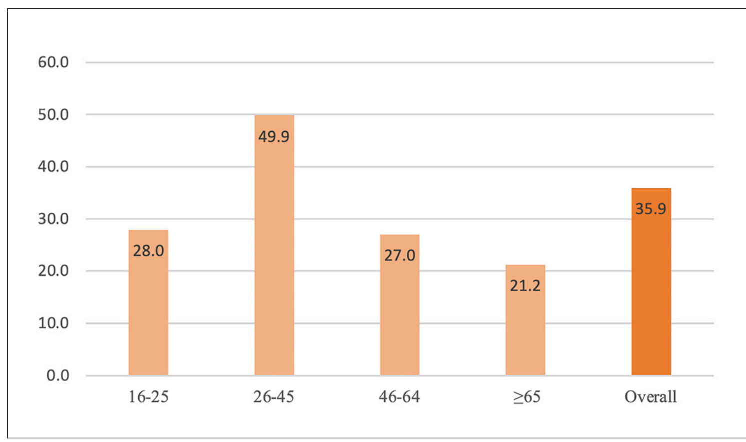

Figure 4. The overall weighted prevalence of glycemic control and across age groups among individuals with diabetes.

Table 4. Frequency of abnormal HA1C tests (>7%).

Morbidity	Controlled	(1-2)	(3-5)	(>5)
None	56.14	42.42	0.8	0.64
One	48.86	27.46	10	13.68
Two	46.56	31.15	12.37	9.92
>Three	22	22.42	20.37	35.22

Total are 100%.

out-of-pocket costs.[24] Case management strategies to address the health of individuals with multimorbidity are needed to improve population health and reduce healthcare costs.

Our finding revealed a high prevalence of obesity that represents a major threat to the health system and Saudi Arabia in general. A recent study estimated that nearly two million obesity-attributed diabetes cases would be diagnosed in Saudi Arabia in the next 20 years.[25] The latest estimate, from a nationally representative sample, estimated the levels of overweight and obesity to be 38% and 20%, respectively.[13] These estimates are lower than the one presented in our study, suggesting a higher prevalence in our population. Obesity is associated with many NCDs and adds to the Kingdom's existing negative impact on population health. A recent local study found that obesity was associated with a two-fold increase in diabetes and hypertension.[7]

Our study revealed that mental conditions prevalence is lower than the currently used national estimate of 20.2%.[26] While it has been found that mental health conditions are underestimated globally, these differences may be due to differences in the catchment population, or difficulty in healthcare access.[27] To reduce the likelihood of missing patients, more screening programs are needed to ensure patients receive the care required.

Our results suggest that the prevalence of injuries is higher among young individuals, similar to the findings in earlier literature.[28] Even when injuries are not fatal, they may cause significant disabilities, affecting population health and healthcare costs.[29] Traffic crashes are the leading cause of injury, fatality, and disability in Saudi Arabia.[30] A recent study estimated that as high as 70% of those who sustain non-fatal traffic crash injuries require long-term rehabilitation.[31] The Kingdom is serious about improving traffic safety, which is why it is one of the main objectives of the health transformation program.[32] Prevention programs may use our findings to guide prioritizing and implementing injury prevention initiatives.

Our findings have major implications for future planning to improve population health. Several steps must be taken to address these conditions as part of population health management. First, there is an imperative to invest in primary care to mitigate the impact of NCDs on population health and health system resources and to reduce the likelihood of adverse health outcomes for the population.[33] This investment could include Chronic Disease Management Programs and clinical guidelines, among other modalities via primary care to improve the management of NCDs and population health.[34]

Second, there is a need to invest in addressing the social determinant of health which influence as much as 80% of health outcomes.[35] There is a need for a comprehensive assessment of social determinants of health in Saudi Arabia and a multi-sectorial approach to their effective management.

Third, there is an opportunity to leverage technology and digital health to enable population health management by identifying social determinants of health, monitoring, and providing appropriate interventions throughout life. Along the care continuum, engaging individuals in managing their health and chronic conditions, establishing disease registries, and regularly monitoring and evaluating the impact of policy interventions. The availability of comprehensive digital data on individuals will make possible the use of advanced data analytics and predictive modeling to identify individuals at risk of chronic conditions and intervene early.

Fourth, continuous monitoring of disease prevalence is crucial to better understand population health, estimate demand on the health system, and project future healthcare expenditures. In the United States, around 90% of the nation's 3.5 trillion annual healthcare expenses are used to manage people with chronic disease or mental health conditions.[36] Therefore, further efforts to screen, diagnose and identify risk profile beneficiaries are needed to activate integrated care pathways to improve health outcomes and reduce health expenditures.[37]

There are three limitations to our study. First, the data is limited to one organization and does not represent the entire Saudi Arabian population. Second, as previously discussed, some individuals with diabetes or hypertension may have never visited any healthcare facilities, thus unaware of the diagnosis. Earlier studies estimated that 57% of diabetic or hypertensive people are unaware of the disease.[12,38] Hence, our findings may underestimate the true prevalence of the conditions studied.

Third, the lack of data on physical activity and diets limits our ability to understand the population's healthy lifestyle, affecting health outcomes. Previous studies from the same underlying population suggest that 73.1% never include moderate exercise in their daily routine.[39] To understand and enhance disease management, other clinical data need to be included in future studies, such as medication adherence and adverse events.

Despite the limitations, there are several strengths of this study. First, we used a population-based sample using minimal additional resources for data collection, setting the base for future and sustainable large-scale studies. Second, we used several objective markers to identify patients with diabetes and hypertension, including machine learning, which improved the reliability of our estimates. Third, we combined various estimates of disease prevalence to understand those with multiple comorbidities and set the stage for future interventions among high-risk groups.

In conclusion, this study indicates that the prevalence of leading NCDs in Saudi Arabia is higher than previously reported in national estimates. Utilizing EHR for monitoring disease prevalence is both feasible and valuable, as regular analysis may reveal trends and changes over time. There is a pressing need for health policies and interventions to address the high levels of multimorbidity, which negatively affect glycemic control. Furthermore, increased investment in assessing the prevalence of major NCDs is essential to accurately estimate their burden and to formulate effective policies for population health management.

Author contributions

SAG, SAQ, and OA conceived the study idea and design. MB, SASH, and AJ reviewed and edited the result section. AJ performed the sensitivity analysis. SAG performed the analysis and wrote the methods section. SASH and MB wrote the result section. RA reviewed and edited the discussion section. All authors reviewed and edited the final manuscript.

REFERENCES

1. The Royal Embassy of Saudi Arabia. Saudi Arabia Political , Economic & Social Development [Internet]. 2017. Available from: https://www.saudiembassy.net/sites/default/files/WhitePaper_Development_May2017.pdf
2. Saudi Arabia Vision 2030, Healthcare transformation [Internet]. 2016 [cited 2021 Jan 4]. Available from: https://www.vision2030.gov.sa/v2030/vrps/hstp/
3. Reka H, Almagrabi A, Alghamdi S. Value-Based Health Care In Saudi Health Insurance Market [Internet]. Riyadh, Saudi Arabia; 2022. Available from: https://chi.gov.sa/ResearchLibrary/CCHI Population Health WP.pdf
4. Tyrovolas S, El Bcheraoui C, Alghnam SA, Alhabib KF, Almadi MAH, Al-Raddadi RM, et al. The burden of disease in Saudi Arabia 1990–2017: results from the Global Burden of Disease Study 2017. Lancet Planet Heal. 2020 May;4(5):e195–208.
5. Global Burden of Disease Compare Tool [Internet]. 2019 [cited 2021 Sep 10]. Available from: https://vizhub.healthdata.org/gbd-compare/
6. Vayena E, Dzenowagis J, Brownstein JS, Sheikh A. Policy implications of big data in the health sector. Vol. 96, Bulletin of the World Health Organization. World Health Organization; 2018. p. 66–8.
7. Alghnam S, Alessy SA, Bosaad M, Alzahrani S, Al Alwan II, Alqarni A, et al. The Association between Obesity and Chronic Conditions: Results from a Large Electronic Health Records System in Saudi Arabia. Int J Environ Res Public Health [Internet]. 2021;18(23):1–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34886087
8. Alhabib KF, Batais MA, Almigbal TH, Alshamiri MQ, Altaradi H, Rangarajan S, et al. Demographic, behavioral, and cardiovascular disease risk factors in the Saudi population: Results from the Prospective Urban Rural Epidemiology study (PURE-Saudi). BMC Public Health. 2020 Aug;20(1):1213.
9. Alotaibi A, Perry L, Gholizadeh L, Al-Ganmi A. Incidence and prevalence rates of diabetes mellitus in Saudi Arabia: An overview. J Epidemiol Glob Health. 2017;7(4):211–8.
10. Hewitt AM, Mascari JL, Wagner SL. Population health management : strategies, tools, applications, and outcomes. 300 p.
11. Al-Ghamdi S, Shubair MM, Aldiab A, Al-Zahrani JM, Aldossari KK, Househ M, et al. Prevalence of overweight and obesity based on the body mass index; A cross-sectional study in Alkharj, Saudi Arabia. Lipids Health Dis. 2018;17(1):1–8.
12. El Bcheraoui C, Basulaiman M, Tuffaha M, Daoud F, Robinson M, Jaber S, et al. Status of the diabetes epidemic in the Kingdom of Saudi Arabia, 2013. Int J Public Health [Internet]. 2014 Dec [cited 2021 Oct 25];59(6):1011–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25292457
13. T A. World Health Survey of Saudi Arabia. Riyadh, Saudi Arabia; 2019.
14. Marcus ME, Ebert C, Geldsetzer P, Theilmann M, Bicaba BW, Andall-Brereton G, et al. Unmet need for hypercholesterolemia care in 35 low- And middle-income countries: A cross-sectional study of nationally representative surveys. PLoS Med. 2021;18(10):1–20.
15. Perlman SE, McVeigh KH, Thorpe LE, Jacobson L, Greene CM, Gwynn RC. Innovations in population health surveillance: Using electronic health records for chronic disease surveillance. Am J Public Health. 2017;107(6):853–7.
16. Patel P, Macerollo A. Diabetes mellitus: Diagnosis and screening. Am Fam Physician. 2010;81(7):863–70.
17. Bailey MK, Weiss AJ, Barrett ML, Jiang HJ. Characteristics of 30-Day All-Cause Hospital Readmissions, 2010–2016: Statistical Brief #248. Healthc Cost Util Proj Stat Briefs [Internet]. 2006;(December 2015):2009–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30896914
18. Marthias T, Anindya K, Ng N, McPake B, Atun R, Arfyanto H, et al. Impact of non-communicable disease multimorbidity on health service use, catastrophic health expenditure and productivity loss in Indonesia: A population-based panel data analysis study. BMJ Open. 2021;11(2):1–13.
19. Abera RG, Demesse ES, Boko WD. Evaluation of glycemic control and related factors among outpatients with type 2 diabetes at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia: a cross-sectional study. BMC Endocr Disord [Internet]. 2022;22(1):1–11. Available from: https://doi.org/10.1186/s12902-022-00974-z
20. Yao L, Mao C, Luo Y. Clinical text classification with rule-based features and knowledge-guided convolutional neural networks. BMC Med Inform Decis Mak. 2019;19(Suppl 3).
21. Deville J-C, Sarndal C-E. Calibration Estimators in Survey Sampling. J Am Stat Assoc [Internet]. 1992 Jun [cited 2021 Oct 24];87(418):376. Available from: https://www.jstor.org/stable/2290268?origin=crossref
22. Alramadan MJ, Magliano DJ, Almigbal TH, Batais MA, Afroz A, Alramadhan HJ, et al. Glycaemic control for people with type 2 diabetes in Saudi Arabia - an urgent need for a review of management plan. BMC Endocr Disord [Internet]. 2018 Sep 10;18(1):62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30200959
23. Atun R. Transitioning health systems for multimorbidity. Lancet [Internet]. 2015;386(9995):721–2. Available from: http://dx.doi.org/10.1016/S0140-6736(14)62254-6
24. Zhao Y, Atun R, Anindya K, McPake B, Marthias T, Pan T, et al. Medical costs and out-of-pocket expenditures associated with multimorbidity in China: quantile regression analysis. BMJ Glob Heal [Internet]. 2021 Feb [cited 2021 Dec 12];6(2). Available from: http://www.ncbi.nlm.nih.gov/pubmed/33632770
25. Coker T, Saxton J, Retat L, Alswat K, Alghnam S, Al-Raddadi RM, et al. The future health and economic burden of obesity-attributable type 2 diabetes and liver disease among the working-age population in Saudi Arabia. Gc VS, editor. PLoS One [Internet]. 2022 Jul 14;17(7):e0271108. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emexb&NEWS=N&AN=636801029
26. Altwaijri YA, Al-Habeeb A, Al-Subaie AS, Bilal L, Al-Desouki M, Shahab MK, et al. Twelve-month prevalence and severity of mental disorders in the Saudi National Mental Health Survey. Int J Methods Psychiatr Res. 2020;29(3):1–15.
27. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. eClinicalMedicine. 2022;54.
28. Alghnam S, Alkelya M, Al-Bedah K, Al-Enazi S. Burden of traumatic injuries in Saudi Arabia: lessons from a major trauma registry in Riyadh, Saudi Arabia. Ann Saudi Med. 2014 Jul;34(4):291–6.
29. Alghnam S, Alkelya M, Aldahnim M, Aljerian N. Healthcare costs of road injuries in Saudi Arabia : A quantile regression analysis. Accid Anal Prev. 2021;159(July 2020):106266.
30. Al-Naami MY, Arafah MA, Al-Ibrahim FS. Trauma care systems in Saudi Arabia: an agenda for action. Ann Saudi Med [Internet]. 30(1):50–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2850182%7B&%7Dtool=pmcentrez%7B&%7Drendertype=abstract
31. Alghnam S, Alghamdi M, Alzahrani S, Alzomai S, Alghannam A, Albabtain I, et al. The prevalence of long-term rehabilitation following motor-vehicle crashes in Saudi Arabia: a multicenter study. BMC Musculoskelet Disord [Internet]. 2022;23(1):1–7. Available from: https://doi.org/10.1186/s12891-022-05153-8
32. Alsofayan Y, Alkhorisi A, Alghnam S, Almalki H, Alsaihani M, Almazroa M, et al. Pursuing Health Sector Transformation Plan, Saudi Vision 2030: Establishing a Trauma Epidemiology Center to Reduce Road Traffic Injuries in Saudi Arabia. Saudi J Emerg Med. 2022;3(1):1.
33. Goldberg DG, Feng LB, Kuzel A. The Role of Primary Care Practices in Advancing Population Health. J Ambul Care Manage [Internet]. 2016 [cited 2021 Jun 19];39(1):87–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26650749
34. Krumholz HM, Currie PM, Riegel B, Phillips CO, Peterson ED, Smith R, et al. A taxonomy for disease management: a scientific statement from the American Heart Association Disease Management Taxonomy Writing Group. Circulation [Internet]. 2006 Sep 26 [cited 2021 May 15];114(13):1432–45. Available from: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.106.177322
35. Hood CM, Gennuso KP, Swain GR, Catlin BB. County Health Rankings: Relationships Between Determinant Factors and Health Outcomes. Am J Prev Med [Internet]. 2016 Feb [cited 2021 May 15];50(2):129–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26526164
36. Health and Economic Costs of Chronic Diseases [Internet]. CDC. 2022. Available from: https://www.cdc.gov/chronicdisease/about/costs/index.htm
37. Sivasampu S, Teh XR, Lim YMF, Ong SM, Ang SH, Husin M, et al. Study protocol on Enhanced Primary Healthcare (EnPHC) interventions: a quasi-experimental controlled study on diabetes and hypertension management in primary healthcare clinics. Prim Health Care Res Dev [Internet]. 2020;21:e27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32787978
38. El Bcheraoui C, Memish ZA, Tuffaha M, Daoud F, Robinson M, Jaber S, et al. Hypertension and Its Associated Risk Factors in the Kingdom of Saudi Arabia, 2013: A National Survey. Int J Hypertens. 2014;2014.
39. Alghnam S, Alyabsi M, Aburas A, Alqahtani T, Bajowaiber M, Alghamdi A, et al. Predictors of Seatbelt Use Among Saudi Adults: Results From the National Biobank Project. Front Public Heal. 2020;8(October):1–10.