The Causal Relationship between Type 2 Diabetes Mellitus and Ovarian Cancer: Two-Sample Mendelian Randomization

Langat Kipkirui Victor¹*, Reuben Cheruiyot Lang’at¹ and Ayubu Anapapa Okango²

¹University of Kabianga, Kenya.
²Murang’a University of Technology, Kenya.

ABSTRACT

Aims: This research aimed at determining the causal relationship between type 2 diabetes mellitus (T2DM) and ovarian cancer using two-sample Mendelian randomization technique. This is because there is an assumption that type 2 diabetes mellitus (T2DM) has a causal relationship with ovarian cancer due to the alarming rising incidence statistics.

Study design: This study used a two-sample Mendelian Randomization (MR) design to undertake the causal relationship investigation. Mendelian randomization technique uses genetic variants as instrumental variables, which undergo random allocation at conception and are non-modifiable. This makes it not to be affected by confounding factors and reverse causation. The MR techniques employed are MR-Egger and Inverse Variance Weighted (IVW.)

Data sources: The outcome (ovarian cancer) summary statistics was retrieved from Ovarian Cancer Association Consortium (OCAC), which has 66,450 samples (number of cases=25,509, number of controls=40,941) of European population. The exposure (T2DM) summary genetic data
came from DIAGRAM plus Metabochip consortium which involved approximately 149,821 samples (number of cases=34,840, number of controls=114,981) of mixed population.

**Results:** The study indicated that there was no evidence of causal relationship between T2DM and ovarian cancer (MR-Egger: b = -0.0476, se = 0.0619, p-value = 0.4479, IVW: b = -0.0165, se = 0.0257, p-value = 0.5217). The odds ratios indicated that the two-sample Mendelian randomization had the power to detect 0.0464 and 0.0164 decrease in variability per 1 SD for MR-Egger and IVW respectively (MR-Egger: OR = 0.9536, CI: 0.8447, 1.0765, IVW: OR = 0.9836, CI: 0.9352, 1.0345).

**Conclusion:** This approach alleviated the usual problem of reverse causation and confounding factors hence depicting clearly that there is no causal relationship between T2DM and Ovarian cancer.

**Keywords:** Type 2 diabetes mellitus; ovarian cancer; mendelian randomization; inverse variance weighted; MR-Egger; causal relationship.

1. INTRODUCTION

People have noticed the rising morbidity and mortality cases associated with cancers and are now blaming the doctors for the late diagnosis of malignancies and the government for lack of efficient facilities to fight this menace [1]. Many of them neglect how their health condition may be contributing to the development of cancer [2, 3]. Ovarian cancer is one of the deadly malignancy due to its late diagnosis and women with type 2 diabetes mellitus (T2DM) have low survival rates from it [4].

Diabetes is one of the metabolic disturbance conditions, which can occur when there is underproduction of insulin, or the body does not effectively use the hormone. Insulin is a hormone that helps significantly in the regulation of blood sugar [5]. The World Health Organization (WHO) reported that diabetes cases have exponentially risen from about 108 million in 1980 to approximately 422 million in 2014 [6]. This situation reflects that the global prevalence had also changed from 4.7% in 1980 to about 8.5% in 2014. The WHO statistics show that in 2012, about 2 million deaths had an association with high blood glucose. This insight depicts that T2DM is one of the menaces in the societies affecting approximately 3.0% to 4.0% of the adults [7].

Kibirige et al. [8] indicated that Africa has a diabetes mellitus disease prevalence of approximately 3.1%. This prevalence translates that about 15.9 million adults in Africa are battling with diabetes mellitus. According to Kibirige et al. [8], most people in the African continent are undiagnosed with the disease hence there is an expectation that the incidences will increase by about 156% by 2045. Mercer et al. [9] noted that most African countries are trying to improve the diabetes care programs, which will ensure accessibility, quality, and safety of medications.

The WHO estimated that the diabetes prevalence in Kenya stands at around 3.3% and it is expected to rise to about 4.5% by 2025. Jones [10] stated that in 2010, diabetes mellitus led to 2% of the total deaths. The Kenyan government have stepped up in helping the people with diabetes by subsidizing the prices of insulin. However, the insulin supply usually runs out and there is mismanagement of funds directed to fighting this menace.

Cancer refers to a collection of associated conditions that lead to abnormal cell growth that has the possibility of spreading [11]. According to 2018 global cancer statistics, there were about 300,000 new ovarian cancer incidences recorded [12]. The 2018 GLOBOCAN estimates indicated that ovarian cancer is the eighth most prevalent malignancy among women globally. Momenimovahed et al. [13] noted that ovarian cancer accounts for about 3.4% of all malignancies in women using GLOBOCAN 295,414 cases. They further stated that approximately 184,799 deaths had an association with ovarian cancer, accounting for about 4.4% of cancer-related demise in 2018. Most of the diagnoses of ovarian cancer usually occur in the advanced stages. The late diagnoses account for about two-thirds of the cases; hence the survival rates tend to be low due to lack of effective screening strategies [14]. These low survival rates necessitate the identification of the predisposing factors to reduce the chances of this type of cancer.

Ovarian cancer is ranked second in Africa among the gynecological malignancies [15]. The major obstacle of the management of this disease in Africa is lack of sufficient screening facilities. This
situation hence leads to the late diagnosis of the ovarian cancer. Most African countries are now combating ovarian cancer through escalation of public awareness and making sure that the machines for detection and diagnosis of the diseases are available. They also try to ensure that there is affordability of the treatment for all cancers.

In Kenya, ovarian cancer is in position three among the major causes of deaths from gynecologic tumors [16]. Like other countries, Kenya experiences a challenge in the diagnosis of the ovarian cancer because of the non-specific characteristics and symptoms at its onset. Due to this reason, more than half of the women with ovarian cancer come to know of their status at the advanced stages [16]. The Kenyan government is currently trying to invest on the screening machines and training of the medical personnel with an aim of curbing the cancer menace as a whole.

Many researches indicate that T2DM is one of the significant predisposing factors for most types of malignancies [17, 7]. The reason for this is that T2DM has a relationship with insulin resistance, hyperinsulinemia, and chronic inflammation, which contributes to the development of cancers [18]. Women with T2DM have ovarian steroid hormone, which alters the levels of estrogen, androgen, and progesterone. For instance, ovarian steroid hormone leads to an increase of estrogen and androgen levels while resulting in the reduction of progesterone. This situation, therefore, creates the potential carcinogenic conditions for the ovaries. T2DM tends to increase insulin or insulin like growth-factors 1 (IGF-1) levels, which have a relationship with the development of ovarian cancer [18]. The reason for this action is that higher levels of insulin and IGF-1 intensifies proliferation and slows down apoptosis in the affected cells, hence leading to the advancement of ovarian malignancy.

Several scholars have tried to investigate the causal relationship between T2DM and ovarian cancer using classical epidemiological methods like case-control and cohort studies [14,19]. Some of the reviews suggested that women with T2DM have a high probability of contracting ovarian cancer than their counterparts. In contrast, other studies indicated that there was no sufficient evidence supporting the relationship. For instance, Wang et al. [14] concluded that women with diabetes mellitus have a high probability of becoming victims of ovarian cancer, especially Asians. On the other hand, Urpilainen et al. [19] demonstrated that there is no proof of an association between T2DM and ovarian cancers among women using metformin or oral anti-diabetic medicines.

Observational studies do not eliminate possible confounding factors, for example, the possible confounders in this study are familial history, the mutation of genes, menstrual periods, and the oral contraceptive usage in the analysis [20]. Therefore, there is need to adopt a better technique that do not suffer the setbacks as the observational methods. MR method is such a technique that can be utilized in the determination of the causal relationship between T2DM and ovarian cancer. Sekula et al. [21] noted that MR is an approach that uses genetic variants as instrumental variables to test the causal relationship between the exposure (T2DM) and the outcome (ovarian cancer). The genetic alleles undergo randomized allocation at conception; hence they are free from confounding factors and reverse causality. Notably, there are limited researches in literature that explains the causal relationship between T2DM and ovarian cancer, and in particular those that have used the two-sample MR approach. There is therefore a need to use a robust method to determine whether there is a causal relationship between T2DM and ovarian cancer.

Mendelian randomization is a research model that assists in establishing the causal relationship between a modifiable predisposing factor (exposure) and the outcome. Sheehan et al. [22] noted that it uses instrumental variables, which makes Mendelian randomization paradigm to be recommendable. These instrumental variables mimic the random allocation of genetic variants to the risk factors. This situation, therefore, ensures that confounding factors and reverse causation does not alter the causal analysis.

This study used a two-sample MR method in the determination of the existence or non-existence of the causal relationship between T2DM and ovarian cancer. Zheng et al. [23] noted that two-sample MR enables the researchers to estimate the causal effects in a case where exposure and outcome data are from samples from totally different populations. Lawlor [24] stated that it is not necessary to obtain the genetic data from the same population. The researcher indicated that using two independent populations will help to
sideline ‘winners’ curse’ that would have led to underestimation of the true causal effects if one group of individuals was used. Apart from that, Lawlor [24] indicated that using two samples in the causal relationship analysis reduces the effects of weak instruments hence increasing the probability of obtaining true causal estimates. This concept makes this model suitable for the study because there are two samples; T2DM and ovarian cancer. Zheng et al. [23] indicated that this method is also advantageous because it greatly increases the scope of Mendelian randomization analysis.

The conceptual framework indicates the causal association between the risk factor (X), T2DM and the outcome (Y), ovarian cancer using the instrumental variants (SNPs). This model indicates that the confounders (C) do not alter the genetic variant. Some of the confounders in this study may include familial history, gene mutation, menstrual periods, and the use of oral contraceptives. Mendelian randomization method has three core assumptions that need fulfillment for it to give unbiased results [25]. The assumptions are as follows:

1. The genetic variant should have a strong relationship with the exposure.
2. The genetic variant should be independent of the confounding factors.
3. The genetic variant should only have a relationship with the outcome via the exposure.

Walker et al. [26] noted that SNPs are credible instruments for determining the causal association between an exposure and a disease outcome because of their random allocation at conception and thus free from subsequent alteration from environmental factors. This insight deduces that if the three assumptions hold, the resulting MR effect estimates are not due to confounding and reverse causation.

Assumption one needs biological support indicating that the gene that encodes the exposure biomarker has the selected genetic variant [21]. This assumption is empirically verifiable since the researcher can use the F statistic, odds ratio, risk ratio, and regression coefficient ($r^2$) to estimate the relationship. Assumptions two and three are not empirically verifiable, however, they are testable to some extent. Koellinger and De Vlaming [27] noted that large-scale Genome-Wide Association Studies (GWASs) have led to the discovering of various genetic loci for many risk factors. This discovery has led to the use of many SNPs in determining the causal link between the exposure and the outcome hence addressing assumption two.

2. METHODOLOGY

2.1 Study Design

This study used two-sample Mendelian randomization (IVW and MR-Egger) technique to investigate the causal relationship between T2DM and ovarian cancer. MR-Egger estimate will be equal to the IVW measure if the intercept is zero. Under the weaker assumption (InSIDE), if the sample sizes and SNPs number increase, the MR-Egger method will give a consistent causal estimate. In a scenario where there is a fixed number of instrumental variables, as the sample size increases the MR-Egger estimate is consistent provided that the inverse-variance weights are equal to zero. Burgess and Thomson [28] indicate that MR-Egger is a significant sensitivity method but it may give biased estimates and inflate Type 1 error rate due to the impacts of the outliers and violation of InSIDE assumptions.

![Fig. 1. Two-sample MR conceptual framework showing the causal relationship between Type 2 Diabetes Mellitus and ovarian cancer](image)
2.2 Data Sources

The summary statistics data was retrieved from two different GWAS consortiums. The exposure genetic (T2DM) data was obtained from the DIAGRAM plus Metabochip consortium. This consortium focuses mainly on samples from European people and performing large-scales studies that try to uncover genetic architecture of type 2 diabetes mellitus. The outcome genetic (ovarian cancer) data was obtained from Ovarian Cancer Association Studies (OCAC). This consortium aims at combining ovarian cancer data from wide range of studies and to give an evaluation of the risk factors of the disease.

2.3 Study Population

The type 2 diabetes mellitus (exposure) data was based on a study done by Morris et al., [29] using DIAGRAMplusMetabochip consortium. They used a sample size of 149,821 (ncase=34,840, ncontrol=114,981) of mixed population. The population had both males and females. The ovarian cancer (outcome) data was based on a study done by Phelan et al., [30] using the summary statistics from Ovarian Cancer Association Consortium (OCAC). This study used a sample size of 66450 (ncase=25,509, ncontrol=40,941) of European population.

2.4 Data Analysis

All the tests were performed using “R” statistical software version 4.0.3 (2020-10-10). Besides that, MR-Base was used as an online platform, which provides an interface allowing MR analyses and sensitivity tests to be performed [26]. The exposure variable in this study was T2DM whose GWAS ID is “ieu-a-24”. It was retrieved from DIAGRAMplusMetabochip consortium. On the other hand, the outcome variable was ovarian cancer whose GWAS ID is “ieu-a-1120”. It was retrieved from Ovarian Cancer Association Consortium (OCAC). Other packages required in R during the analysis include: “devtools”, “Two Sample MR”, “digest”, “githubinstall”, and “google Auth R”.

The researcher proceeded to extract the exposure data, type 2 diabetes mellitus and clump them. This move assisted in identifying the independent alleles among the correlated SNPs [26]. The next step was to list all the available outcomes in the MR-Base platform and extract the outcome data, ovarian cancer. In this case, the researcher allowed the use of SNP proxy, which was in LD with the targeted SNP. The researcher defined the minimum r-square to find the SNP proxy to be 0.8. On the other hand, the study assumed that all the alleles are aligned in the forward strand. According to Walker et al., [26], palindromic SNPs refer to a situation where pair of alleles on the forward-strand are the same as those on the reverse strand. This study infers the palindromic SNPs and the maximum minor allele frequency acceptable threshold defined by 0.3.

The next step was to harmonize the exposure and the outcome data. Walker et al., [26] defined harmonization as a way of specifying the effect and other alleles in the same way in both the exposure and outcome data. In this study, interpretation of the forward strand was by use of the allele frequency information. At this point, it was possible to perform the two-sample Mendelian randomization analyses. This gave the results of five different Mendelian randomization methods; MR-Egger, weighted median, IVW, simple mode, and weighted mode. Since this study focused mainly on MR-Egger and IVW methods, the researcher restricted the scatter plot to only depict these two techniques.

3. RESULTS

The research found 39 variants of the exposure variable (T2DM) after clumping the data. Besides that, the study found the proxies for 3 SNPs in the outcome data (ovarian cancer). After harmonizing the study data, 3 SNPs (rs10830963, rs1801282, rs243088) were found to be palindromic with intermediate allele frequencies. Therefore, the study utilized 33 SNPs in the analysis.

It was necessary to perform some of the sensitivity analyses. This action will either support or question the validity of the research. The interpretation of the odds ratios (exponents of beta coefficients) of Mendelian randomization assisted in determining the validity of the causal inference from Mendelian randomization as shown in Table 2. Therefore, this part will give the validity and reliability of the study.

4. DISCUSSION

The prime objective of this research study was to investigate the causal relationship between T2DM and ovarian cancer using two-sample Mendelian randomization. This investigation was
Fig. 2. Scatter plot representing two-sample Mendelian randomization results of the causal relationship between Type 2 Diabetes Mellitus and Ovarian Cancer: The black dots represents each of the SNPs associated with T2DM while the horizontal and the vertical lines depict the standard error of the relationship between type 2 diabetes mellitus and ovarian cancer respectively.

necessary because there are studies that have identified that some of the hormones associated with high blood sugars tend to create carcinogenic conditions hence accelerating the growth of cancer [18]. Apart from that, other scholars have used the observational models like cohort and case control to determine the relationship between T2DM and ovarian cancer but have ended up with conflicting points of view [14,19]. Therefore, this research used two-sample Mendelian randomization, a technique which is not prone to confounding factors and reverse causation unlike the observational studies.

The data indicated that the exposure p-values were less than 0.05, which means that T2DM is strongly associated with the targeted 33 SNPs. Contrary to that, the outcome p-values were greater than 0.05 indicating that ovarian cancer are only associated with the targeted 33 SNPs via the exposure. This situation shows that this study fulfills the assumptions of Mendelian randomization. The F-statistic ($F= 65.269, P=0.000$) for this study was greater than 10, the GWAS standard threshold, as shown in Fig. 1. This reflects that the SNPs used in this study are considered to be strong instrumental variables as suggested by [31].

The causal estimates for the relationship between T2DM and ovarian cancer were obtained through the interpretation of the beta coefficients as shown in Table 1. The causal relationship output indicated that the MR-Egger beta coefficient was -0.048 with se=0.062 and p-value=0.448. On the other hand, the IVW method had a beta coefficient of -0.016 with se=0.026.
### Table 1. The two-sample Mendelian randomization results of the causal relationship between type 2 diabetes mellitus and ovarian cancer

<table>
<thead>
<tr>
<th>Method</th>
<th>nsnp</th>
<th>b</th>
<th>exp(b)</th>
<th>se</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR Egger</td>
<td>33</td>
<td>-0.048</td>
<td>0.954</td>
<td>0.062</td>
<td>0.448</td>
</tr>
<tr>
<td>Weighted median</td>
<td>33</td>
<td>-0.042</td>
<td>0.959</td>
<td>0.036</td>
<td>0.243</td>
</tr>
<tr>
<td>Inverse variance weighted</td>
<td>33</td>
<td>-0.016</td>
<td>0.984</td>
<td>0.026</td>
<td>0.522</td>
</tr>
<tr>
<td>Simple mode</td>
<td>33</td>
<td>-0.031</td>
<td>0.970</td>
<td>0.072</td>
<td>0.674</td>
</tr>
<tr>
<td>Weighted mode</td>
<td>33</td>
<td>-0.041</td>
<td>0.960</td>
<td>0.040</td>
<td>0.323</td>
</tr>
</tbody>
</table>

*nsnp: number of SNPs, b: beta coefficient, exp(b): exponential of beta coefficient, se: standard error, pval: p-value*

### Table 2. The odds ratios results of the two-sample Mendelian randomization of the causal relationship between type 2 diabetes mellitus and ovarian cancer

<table>
<thead>
<tr>
<th>Method</th>
<th>nsnp</th>
<th>b</th>
<th>se</th>
<th>pval</th>
<th>Lower CL</th>
<th>Upper CL</th>
<th>OR</th>
<th>OR Lower CL</th>
<th>OR Upper CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR Egger</td>
<td>33</td>
<td>-0.048</td>
<td>0.062</td>
<td>0.448</td>
<td>-0.169</td>
<td>0.074</td>
<td>0.954</td>
<td>0.845</td>
<td>1.076</td>
</tr>
<tr>
<td>Weighted median</td>
<td>33</td>
<td>-0.042</td>
<td>0.036</td>
<td>0.243</td>
<td>-0.113</td>
<td>0.029</td>
<td>0.959</td>
<td>0.893</td>
<td>1.029</td>
</tr>
<tr>
<td>Inverse variance weighted</td>
<td>33</td>
<td>-0.016</td>
<td>0.026</td>
<td>0.522</td>
<td>-0.067</td>
<td>0.034</td>
<td>0.984</td>
<td>0.935</td>
<td>1.035</td>
</tr>
<tr>
<td>Simple mode</td>
<td>33</td>
<td>-0.031</td>
<td>0.072</td>
<td>0.674</td>
<td>-0.171</td>
<td>0.110</td>
<td>0.970</td>
<td>0.842</td>
<td>1.117</td>
</tr>
<tr>
<td>Weighted mode</td>
<td>33</td>
<td>-0.041</td>
<td>0.040</td>
<td>0.323</td>
<td>-0.120</td>
<td>0.039</td>
<td>0.960</td>
<td>0.887</td>
<td>1.039</td>
</tr>
</tbody>
</table>

*nsnp: number of SNPs, b: beta coefficient, se: standard error, pval: p-value, CL: confidence level, OR: odds ratio*
and p-value= 0.522. It was noticeable that the p-values for these Mendelian randomization methods were greater than 0.05 indicating that there is no sufficient evidence of the causal relationship between T2DM and ovarian cancer. The other three methods (weighted median: b=-0.042, se=0.036, p-value=0.243, simple mode: b=-0.031, se=0.072, p-value=0.674, and weighted mode: b=-0.041, se=0.040, p-value=0.323) gave consistent results, which showed that MR-Egger and IVW techniques are capable of giving reliable results. The results in Table 1 were graphically represented where the effects of the SNPs on exposure (T2DM) were against the effects of the SNPs on the outcome (ovarian cancer) as suggested by Walker et al. [26] as shown in Fig. 2. In Fig. 3, the causal estimates of each SNP used in this study are shown which were generated using the Wald ratio. Apart from that, the graph indicates

Fig. 3. Forest plot displaying the results of single and multi-SNP analyses on the causal relationship between Type 2 Diabetes Mellitus and Ovarian Cancer
Fig. 4. Leave-one-out graph of the causal relationship between Type 2 Diabetes Mellitus and Ovarian Cancer

the multi-SNP causal estimates using the MR-Egger and IVW techniques. This graph shows that there was no discrepancy in the causal estimates displayed. This revelation disapproves the research done by Wang et al. [14] that led to the conclusion that diabetic women, especially, Asians are highly prone to ovarian cancer. On the other hand, the study by Urpilainen et al. [19] indicated that the medication by diabetic women maybe contributing to the growth of ovarian cancer and not the diabetes disease.

The odds ratios helped in indicating the power (sensitivity) of this two-sample Mendelian
randomization study. First, it is necessary to note that taking the exponent of beta coefficient is the same as the odds ratio (OR) as shown in Table 2. The OR for MR-Egger method was 0.9536 with confidence interval (CI) (0.8447, 1.0765). This indicates that the model, MR-Egger was capable of detecting 0.0464 decrease of variability per 1 standard deviation (SD). On the other hand, IVW technique had the OR of 0.9836 with CI (0.9352, 1.0345). This reflects that IVW was able to detect 0.0164 decrease of variability per 1 SD. The other methods gave the OR which were in the same range as those of MR-Egger and IVW (weighted median: OR=0.959, CI: 0.893, 1.029, simple mode: OR=0.970, CI: 0.842, 1.117, weighted mode: OR=0.960, CI: 0.887, 1.039), hence showing that the two-sample Mendelian randomization gives consistent causal inference. The leave-one-out analysis in Fig.4 uses IVW method when excluding one SNP each time [26]. This leave-one-out graph indicates that all the selected SNPs in this study were consistent hence there were no potentially influential SNPs. Therefore, the researcher can conclude that the results of this study were not influenced by a single outlying SNP.

It is noticeable that most of the genetic data were extracted majorly from the participants from the European origin. Therefore, as a limitation of this study, this situation may not guarantee the generalization of the findings in other people from different places of the world whose SNPs were not captured. Therefore, the study recommends the research institutions to invest in getting the genome data from all the regions of the world. This will increase the scope of the genome analysis and improve precision medicine.

5. CONCLUSION

This research study has found out that there is no sufficient evidence of the causal relationship between T2DM and ovarian cancer as shown by the beta coefficients in Table 1. It can be concluded that Mendelian randomization technique is a robust method. This is so because it uses genetic variants (SNPS) that undergo random allocation at conception and are non-modifiable. These properties of MR technique enables it not to be prone to confounding factors and reverse causation. This problems are common in the other techniques such as observational methods. Two-sample Mendelian randomization increases the scope of the study since it uses the genetic data from samples obtained from totally different populations.

CONSENT

Not applicable.

ETHICAL APPROVAL

Not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

7. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JP. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ. 2015; 350:g7607.[CrossRef]
8. Kibirige D, Lumu W, Jones AG, Smeth L, Hattersley AT, Nyirenda MJ. Understanding the manifestation of
23. Zheng J, Baird D, Borges MC, Bowden J, Hemani G, Haycock P, Smith GD. Recent Developments In Mendelian Randomization Studies. Current epidemiology this research study has found out that there is no causal relationship between T2DM and ovarian cancer and thus other causalities may therefore be of interest in investigation. Reports. 2017;4(4):330-34. [CrossRef]
27. Koellinger PD, De Vlaming R. Mendelian randomization: The challenge of
unobserved environmental confounds; 2019. [CrossRef]


Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/70002