Counterfactual Inference Using Ordinary Differential Equations to Assess the Effect of Physical Activity on Type 2 Diabetes Onset

Marta Lenatti[1 [0000−0003−0731−7505]], Marco Zaffalon[2 [0000−0001−8908−1502]], Alessandro Antonucci[2 [0000−0001−7915−2768]], Pierluigi Francesco De Paola[1,3 [0009−0005−5198−0293]], Lea Multerer[2 [0000−0002−4667−9034]], Maurizio Mongelli[1 [0000−0001−6201−6225]], Alessia Paglialonga[1✉ [0000−0002−1341−2560]], and Laura Azzimonti[2 [0000−0003−0565−3505]]

[1] Consiglio Nazionale delle Ricerche (CNR), 00185 Rome, Italy
alessia.paglialonga@cnr.it
[2] IDSIA-SUPSI, 6900, Lugano, Switzerland
[3] Politecnico di Bari, 70126 Bari, Italy

Abstract. This study introduces a counterfactual inference approach for dynamical models based on ordinary differential equations (ODEs). The proposed methodology is applied to a system of ODEs describing blood glucose regulation to assess the long-term effects of personalized physical activity plans on type 2 diabetes progression. The analysis demonstrates how responses to physical activity vary among individuals based on β-cells mass and insulin sensitivity, resulting in diverse sets of counterfactual physical activity plans. This approach can eventually help identify individuals who might benefit from early diabetes prevention measures via personalized physical activity plans.

Keywords: Personalized treatment · Counterfactual inference · ODEs.

1 Introduction

Promoting physical activity is a key strategy in Type 2 Diabetes (T2D) prevention programs [8, 14]. According to the World Health Organization (WHO) 2020 guidelines on physical activity [15] all adults should engage in regular exercise consisting of at least 150 minutes of moderate-intensity aerobic physical activity, or at least 75 minutes of vigorous-intensity aerobic physical activity per week to slow down or even prevent progression to T2D. Although generic guidelines exist and were proved to be effective in general, such recommendations may not be adequate for all subjects. Therefore, personalized physical activity plans are required, based on the subject's characteristics and needs [2, 13]. A promising approach to tailor treatments is offered by causal inference, which naturally addresses what-if analyses by means of counterfactual reasoning [5].

This work presents a counterfactual inference approach for time-dependent settings that relies on casting a system of ordinary differential equations (ODEs)

into the causal framework. The proposed methodology has been applied to analyze the long-term effect of personalized hypothetical physical activity plans on T2D progression, leveraging a recently developed ODEs model of blood glucose regulation [3]. The insights gained from this retrospective approach can inform the design of more precise guidelines tailored for distinct groups of individuals with shared characteristics.

2 Methodology

2.1 System of Ordinary Differential Equations

Systems of ODEs are mathematical models commonly used in biomedicine to describe dynamic phenomena, thus providing a comprehensive representation of their evolution over time. The approach here presented is based on the assumptions that (i) there exists a system of ODEs modeling the phenomenon under analysis, (ii) the solution of the system exists and is unique, and (iii) a numerical solver for the system is available. A system consistent with these assumptions is the model developed by De Paola et al. [3], which integrates and extends two well-established models of glucose and insulin dynamics [**?**,**?**] to characterize short- and long-term effects of physical activity on T2D progression. The model, consisting of 12 ODEs and over 50 parameters, can simulate an average adult subject with no assumptions about age, sex assigned at birth, and fitness level.

To increase computational efficiency, we used an approximated version of the aforementioned model that captures the system's macroscopic dynamics [10] and consists of seven state variables $[G, I, B, S_{\mathrm{I}}, VL, \Sigma, \Gamma]$ with initial conditions $[G_0, I_0, B_0, S_{\mathrm{I}0}, VL_0, \Sigma_0, \Gamma_0]$. Physical activity is modeled as a squared wave function PA that evolves over time and depends on three parameters: the intensity of exercise above the basal level ξ, defined from 0 to 92%, the period ν, in days, between two consecutive sessions, and the duration δ, in minutes, of each training session. PA can be represented in a more compact way using the constant average value $\overline{PA} = \xi\delta/\nu$ [4]. Both the original ODEs model and its approximated version describe the cumulative effect of Interleukin-6 release during physical exercise (VL) that improves insulin sensitivity (S_{I}) and increases the mass of β cells (B). This improvement results in a higher basal insulin concentration (I), reducing basal glucose levels (G) and providing a long-term protective effect through the insulin-glucose feedback mechanism. For further details on state variables and parameters, please refer to [10].

To model subjects with different characteristics and propensities for T2D progression, we sampled three initial conditions and six model parameters within clinically acceptable ranges using predefined step sizes (Table 1). Sobol sensitivity analysis was conducted using the SALib Python library [6, 7] to measure the influence of each variable on the output of interest, i.e., G after five years.

2.2 Structural Causal Model

The proposed causal inference approach is based on the structural causal model (SCM) framework [11], which has been widely adopted to quantify causal rela-

Table 1: Initial conditions and model parameters that have been varied.

Variable	Unit	Description	Range	Step
G_0	mg/dl	Plasma glucose concentration at t=0	70-120	10
I_0	μU/ml	Serum insulin concentration at t=0	5-20	5
B_0	mg	β-cells mass at t=0	800-1300	100
W	kg	Weight	50-150	20
τ_{S_I}	days	Insulin sensitivity time constant	90-330	60
$S_\mathrm{I}^{\mathrm{target}}$	ml/μU/days	Insulin sensitivity target value	0.18-0.68	0.10
ξ	%	Intensity of physical activity	10-70	10
$1/\nu$	times/week	Weekly frequency of physical activity	1-3	1
δ	minutes/session	Duration of a physical activity session	30-90	30

tionships from observational data. Within this framework, the causal relationships between variables are expressed through deterministic structural equations. The ODEs model in [10] provides deterministic (differential) relations between variables and allows for the derivation of a dynamic causal model M that satisfies the SCM theory, as demonstrated in [1]. The time-dependent causal model M is associated with a directed graph $\mathcal{G}(M)$ characterized by nodes and edges. The set of nodes comprises exogenous and endogenous variables, which are represented by U and V, respectively. Each directed edge represents a causal relationship between a couple of nodes. In the examined case, the triplet $\langle U, V, F\rangle$ characterizing M consists of: $U = \{G_{\mathrm{obs}}, I_0, B_0, W, S_\mathrm{I}^{\mathrm{target}}, \tau_{S_\mathrm{I}}, v\}$ representing patient-specific conditions or observations (e.g., G_{obs}); $V = \{G, I, B, \Sigma, \Gamma, S_\mathrm{I}, VL, PA\}$ representing dynamic processes that are determined by other variables in $U \cup V$; $F = \{f_1, ..., f_m\}$, functions derived from ODEs, mapping $U \times (V \backslash V_j)$ to a specific endogenous variable V_j, $j = 1, \ldots, m$. Figure 1 presents the graph $\mathcal{G}(M)$ of the selected ODEs model, with the exogenous nodes displayed in gray and the endogenous nodes, indicating both the state variable and its first-order derivative, shown in white. In such a setup, if $\mathcal{U}$ denotes the set of possible values for U, an instantiation $u \in \mathcal{U}$ of the exogenous variables deterministically induces the values of all the endogenous variables and therefore represents a specific individual. However, some of the exogenous variables may not be determined in real-world settings, for instance because they cannot be easily measured (e.g., B_0, τ_{S_I}, $S_\mathrm{I}^{\mathrm{target}}$) or because they represent latent factors (e.g., v which indicates the willingness of a subject to follow the prescribed physical activity plan). As a result, we may only have partial knowledge of an individual. If E denotes a set of observable variables (i.e., $E = \{G_{\mathrm{obs}}, W, PA\}$) then, for each partial evidence $E = e$, $\mathcal{U}_e \subseteq \mathcal{U}$ represents the set of exogenous nodes values compatible with e, while u_e represents one of its elements [12].

2.3 Counterfactual Inference

Counterfactual inference examines what-if scenarios to understand the potential impact of certain hypothetical actions on an outcome of interest [11]. Such an analysis can inform decisions in both probabilistic and deterministic settings. Let X and Y be two variables in V, with X being the variable to manipulate (cause) and Y being the outcome (effect) that we want to monitor under evidence e.

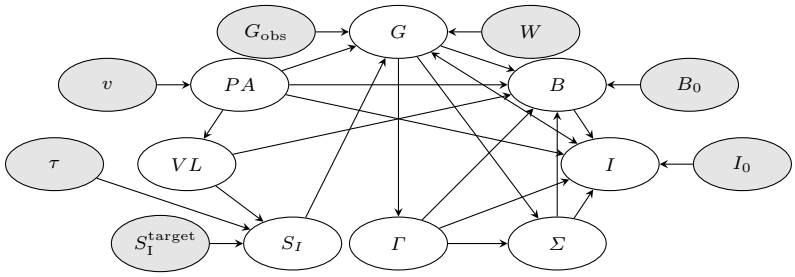


Fig. 1: Causal graph $\mathcal{G}(M)$ of the ODEs model in [10].

The counterfactual query $Y_x(u_e)$ represents the hypothetical value that Y would have taken if X had been x, based on a specific instantiation u_e of the exogenous variables compatible with e [12]. When dealing with partial evidence, as in the current analysis, the problem is non-identifiable. Therefore, the query must be evaluated for each $u_e \in \mathcal{U}_e$, thus yielding a set of solutions [16].

Let $X = PA$, $Y = G$, and $\mathcal{PA}$ be the set of possible physical activity plans. In this study, we assumed that given the current (factual) physical activity plan $pa \in \mathcal{PA}$, glucose at $t = t_{\text{end}}$ exceeds the diabetic threshold, i.e., $G|_{t=t_{\text{end}}} \geq thr = 125$ mg/dl. We were interested in evaluating the counterfactual query $G_{pa^*}(u_e)|_{t=t_{\text{end}}}$, which represents *the value that G would have obtained at $t = t_{\text{end}}$ had pa been different, namely $pa^* \in \mathcal{PA}\backslash\{pa\}$* and in identifying all physical activity plans that could have prevented the subject from progressing from normoglycemia or prediabetes to T2D over the five-year period, according to the model. To compute the counterfactual query, first we supposed to observe $e = \{\boldsymbol{g}_{\mathbf{obs}}, w, pa\}$, where $\boldsymbol{g}_{\mathbf{obs}} = [g_0, \ldots, g_{n_{\text{obs}}-1}]$ comprises one or more glucose measurements at instant $t_0, \ldots, t_{n_{\text{obs}}-1}$, such that g_0 is the glucose value at $t = 0$ and $t_{n_{\text{obs}}-1} < t_{\text{end}}$ with $t_{\text{end}} = 5$ years, whereas w and pa are fixed during the observation period. Then, we followed these steps:

1. *Abduction.* Determine compatible sets of values for the exogenous nodes, given the evidence. Observing e and $G|_{t=t_{\text{end}}} \geq thr$ reduces the set of possible values of U from $\mathcal{U}$ to $\mathcal{U}_e$, with $\mathcal{U}_e = \{u \in \mathcal{U} \mid E = e, G|_{t=t_{\text{end}}} \geq thr\}$, $\mathcal{U}_e \subseteq \mathcal{U}$.
2. For each compatible configuration of the exogenous nodes $u_e \in \mathcal{U}_e$ and for each hypothetical physical activity plan $pa^* \in \mathcal{PA}\backslash\{pa\}$:
   - *Action:* perform the intervention by forcing $PA = pa^*$.
   - *Prediction:* compute the potential response of G to action, i.e., $G_{pa^*}(u_e)$.

   Once we have found the glucose trajectories that align with the observed evidence for each physical activity plan $pa^* \in \mathcal{PA}\backslash\{pa\}$, we determine the set of counterfactual physical activity plans as the hypothetical values of PA that could have prevented T2D at t_{end}, i.e.,

$$\widetilde{\mathcal{PA}}_{u_e} = \{pa^* \in \mathcal{PA}\backslash\{pa\} \mid G_{pa^*}(u_e)|_{t=t_{\text{end}}} < thr\}, \tag{1}$$

   and the least demanding plan $pa^{*\min}_{u_e}$, characterized by the average value

$$\overline{pa}^{*\min}_{u_e} = \min_{pa^* \in \widetilde{\mathcal{PA}}_{u_e}} \overline{pa}^*. \tag{2}$$

3. Lastly, we consider the most conservative physical activity plan with a preventive factor $pa_e^{*\max}$, characterized by the average value

$$\overline{pa}_e^{*\max} = \arg\max_{u_e \in \mathcal{U}_e} \overline{pa}_{u_e}^{*\min}. \quad (3)$$

The optimization in Eq. (2) and (3) might be extremely challenging: besides exploring a multidimensional domain, it requires an analytical solution of the ODEs model in closed form. Since this solution is not available in the system under study, the counterfactual query is approximated using a numerical solution. First, we generated a dataset $\mathcal{D}$ that covered a broad range of simulation scenarios, considering the conditions in Table 1. Then, we performed counterfactual inference as described in **Algorithm 1**. It should be noted that the quantities retrieved by the application of **Algorithm 1** serve as a discrete approximation of the real quantities, as a consequence of the finite size of $\mathcal{D}$. The set of simulated individuals in $\mathcal{D}$ that are compatible with the evidence (factual simulations) is stored in $\mathcal{D}_e$. The computational complexity of the algorithm for each e is $O(|\mathcal{U}_e|)$ if $n_{\text{obs}} = 1$ and $O(|\mathcal{U}_e| \cdot |\mathcal{PA} \backslash \{pa\}|) \approx O(|\mathcal{U}_e| \cdot |\mathcal{PA}|)$ otherwise. Related codes will be made publicly available upon acceptance of this manuscript.

Algorithm 1: Counterfactual Inference

Input: A dataset $\mathcal{D}$ of model simulations, a partial evidence e
Output: The value $\overline{pa}_e^{*\max}$

$\mathcal{D}_e \leftarrow$ all rows in $\mathcal{D}$ where $E = e$ and $G|_{t=t_{\text{end}}} \geq thr$;
$\mathcal{U}_e \leftarrow$ all combinations $u_e \in \mathcal{D}_e$;
For each $u_e \in \mathcal{U}_e$ **do:**
  if $n_{\text{obs}} = 1$ **then**
    $\widetilde{\mathcal{PA}}_{u_e} \leftarrow$ all $pa^* \in \mathcal{PA} \backslash \{pa\}$ in $\mathcal{D} \backslash \mathcal{D}_e$ compatible with u_e, e, and with $G|_{t=t_{\text{end}}} < thr$, as defined by Eq. (1);
  else
    New initial conditions $IC' \leftarrow$ current output at time $t_{n_{\text{obs}}-1}$;
    For each $pa^* \in \mathcal{PA} \backslash \{pa\}$ **do:**
      sim $\leftarrow$ `RunModelSimulation`(IC', u_e, $t_{\text{end}} - t_{n_{\text{obs}}-1}$);
      if *sim has* $G|_{t=t_{\text{end}}} < thr$ **then**
        Save pa^* in $\widetilde{\mathcal{PA}}_{u_e}$;
    end
    $\overline{pa}_{u_e}^{*\min} \leftarrow$ minimum value within $\widetilde{\mathcal{PA}}_{u_e}$, defined by Eq. (2);
  end
end
$\overline{pa}_e^{*\max} \leftarrow$ maximum value of $\overline{pa}_{u_e}^{*\min}$ across all u_e by Eq. (3);

3 Results

3.1 Simulation Phase

A dataset $\mathcal{D}$ composed of 1,814,400 five-year simulations of the ODEs model [10] was generated to model diverse subjects with variations in glucose-insulin

metabolism, accounting for all variable combinations in Table 1. Simulations were conducted on a LENOVO workstation equipped with an Intel Core i7-10700K CPU (8 cores, 16 threads, 2.9 GHz) and 32 GB of RAM operating on Windows 10 Pro. Each five-year simulation took approximately 0.025 seconds.

In the dataset, 6.85% of the subjects developed T2D during the simulation period. Of these, 94.1% engaged in physical activity that is insufficient according to the WHO guidelines [15]. Normoglycemic or prediabetic subjects at the beginning of the simulation were never predicted to progress to T2D in five years if their initial β-cell mass exceeded 1200 mg and/or if $S_{\mathrm{I}}^{\mathrm{target}} \geq 0.28$ ml/μU/d.

The variables that most significantly influenced the output according to Sobol sensitivity analysis were $S_{\mathrm{I}}^{\mathrm{target}}$ (0.91), B_0 (0.70), $\overline{PA}$ (0.44), and $\tau_{S_{\mathrm{I}}}$ (0.33). W presented a lower impact (2.33×10^{-3}) whereas G_0 and I_0 had almost no impact on the output (total sensitivity $< 1 \times 10^{-6}$). Only a small part of the total variability (e.g., 0.18 for $S_{\mathrm{I}}^{\mathrm{target}}$) was due to first-order terms (i.e., due to changes in a single input parameter), while most of the total variability was due to interaction terms (i.e., simultaneous changes in more than one input parameter).

3.2 Counterfactual Inference Phase

During the inference phase, we hypothesized a set of potential observations e ($n_{\mathrm{obs}} = 1$) and ranges of $G|_{t=t_{\mathrm{end}}}$, resulting from all possible combinations of the following values: $g_0 \in \{70, 120\}$, $w \in \{50, 70, 90, 110, 130, 150\}$, and 9 fixed physical activity plans with $1/\nu \in \{1, 2, 3\}$, $\xi \in \{30, 50, 70\}$, $\delta = 60$ minutes and $G|_{t=t_{\mathrm{end}}} \in \{(125, 150), (150, 175), (175, 200), (200, 600)\}$. Out of 432 combinations, 280 had factual simulations in the specified range of $G|_{t=t_{\mathrm{end}}}$, with a median of 16 compatible simulated individuals for each observation e (min=4, max=44). These simulations, in turn, yielded a median of 11 counterfactual physical activities each (min=0; max=51). Additionally, 108 observations presented at least one factual simulation without a corresponding counterfactual physical activity plan. These factual simulations represented subjects whose progression to T2D could not be prevented by modifying physical activity alone, as their glucose metabolism was already compromised. They were characterized by a rapid decay of insulin sensitivity ($\tau_{S_{\mathrm{I}}} = 90$ days, $S_{\mathrm{I}}^{\mathrm{target}} = 0.18$ ml/μU/d) and a reduced β-cells mass ($B_0 = 800$ or 900 mg), resulting in hyperglycemia at the end of the observation period ($G|_{t=t_{\mathrm{end}}}$ between 200 and 600 mg/dl).

Table 2 describes the effect of the factual pa, B_0 and $\tau_{S_{\mathrm{I}}}$ on the least demanding counterfactual plan, $\overline{pa}_{u_e}^{*\min}$, across the set of factual simulations.

Table 2: Effect of factual pa, B_0 and $\tau_{S_{\mathrm{I}}}$ on $\overline{pa}_{u_e}^{*\min}$ (median and IQR).

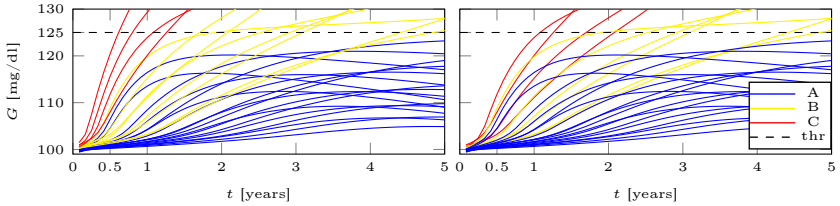
Factual *pa*			***B*$_0$**		***τ*$_{S_{\mathrm{I}}}$**	
2/week 50%	**3/week 50%**	**3/week 70%**	**800 mg**	**1100 mg**	**90 days**	**330 days**
1.04 (0.63)	1.36 (0.59)	1.63 (0.44)	1.36 (0.86)	0.81 (0.10)	1.63 (1.07)	0.71 (0.30)

As shown in the table, both the frequency and intensity of the factual pa

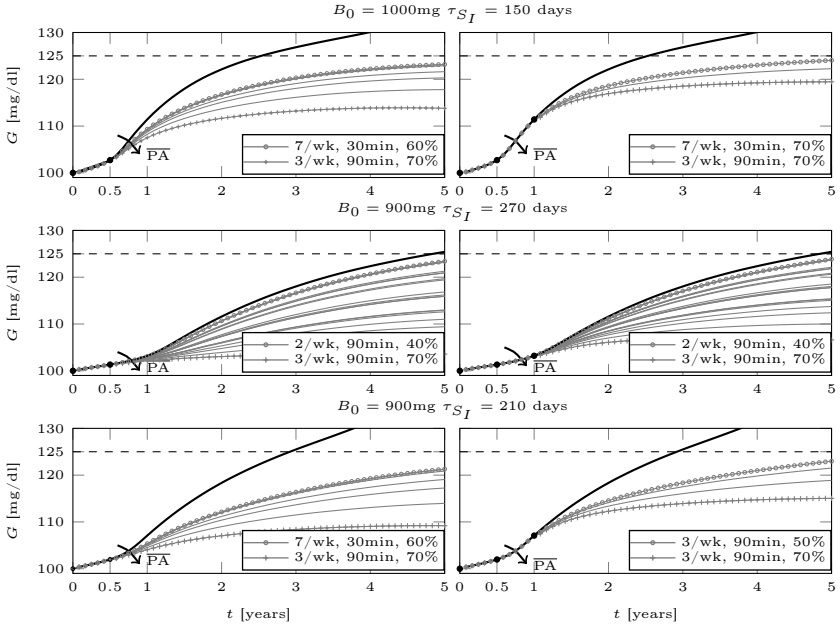
influence $\overline{pa}^{*\min}_{u_e}$, with a greater demand when the intensity and frequency currently observed are higher. Moreover, the lower the β-cells mass, the greater the amount of physical activity (higher $\overline{pa}^{*\min}_{u_e}$) hypothetically required to prevent T2D onset, while lower values of τ_{S_I}, indicating a faster decrease in insulin sensitivity, require a greater physical activity to prevent T2D onset. The analysis of the distribution of $\overline{pa}^{*\max}_{e}$, that is, the plan ensuring T2D prevention across all unknown exogenous variables u_e that are consistent with the observed evidence e, shows that $\overline{pa}^{*\max}_{e}$ heavily depends on the factual pa and on $G|_{t=t_\text{end}}$. However, the calculation of $\overline{pa}^{*\max}_{e}$ as in Eq. (3) uses a conservative approach, which can lead to an over-pessimistic estimate, i.e., always suggesting the most intense physical activity if e is not sufficiently informative. This estimation can be improved by the availability of additional measures of glucose, i.e., $n_\text{obs} > 1$, which allows a better estimation of unknown exogenous parameters and a more specific estimation of $\overline{pa}^{*\max}_{e}$, as illustrated in the followings.

Example: Individual response to WHO-compliant physical activity. The response of different individuals to the same physical activity plan may vary depending on genetic and acquired factors that could, for instance, affect their degree of insulin sensitivity and β-cells functionality. Indeed, an individual who is prediabetic and follows the WHO guidelines [15] could still progress to T2D according to the model. As an example, we simulated individuals with a weight of 70 kg who engaged in vigorous-intensity aerobic physical activity for 90 minutes per week, with observed prediabetic measurements at six months ($n_\text{obs} = 2$ with $t_1 = 0.5$ years, left-hand side) and at one year ($n_\text{obs} = 3$ with $t_1 = 0.5$ and $t_2 = 1$ year, right-hand side). Figure 2a shows the set of possible glucose trajectories of such individuals. Some individuals may not develop T2D within the next five years, thus not requiring intervention (Group A in Figure 2a, corresponding to simulations with $B_0 \geq 1100$ mg and $\tau_{S_\text{I}} \geq 150$ days, $B_0 = 1000$ mg and $\tau_{S_\text{I}} \geq 210$ days, or $B_0 = 900$ mg and $\tau_{S_\text{I}} \geq 330$ days). For most of the simulated individuals who develop T2D, disease progression could have been prevented (Group B), whereas in a subset of subjects, modifying physical activity alone would have been insufficient to reverse T2D progression, mainly due to impaired insulin sensitivity and β-cells dysfunction (Group C, corresponding to simulations with $B_0 \leq 1000$ mg and $\tau_{S_\text{I}} = 90$ days or $B_0 = 800$ mg and $\tau_{S_\text{I}} \leq 150$ days). More specifically, Figure 2b illustrates three examples of preventable factual glucose trajectories selected from Group B of Figure 2a (black lines) and the corresponding counterfactual trajectories (gray lines). Given the same evidence, the progression towards T2D is considerably influenced by the degree of insulin sensitivity impairment and β-cells dysfunction, particularly in terms of defects in β-cells mass due to increased apoptosis. The faster a patient is expected to progress to T2D, the more limited is the range of potential counterfactual plans, making it more challenging to prevent the disease through exercise. Furthermore, as anticipated, an elevated $\overline{PA}$ for the counterfactual plan is associated with a reduced glucose value at $t = t_\text{end}$, thereby reducing the likelihood of developing T2D. Lastly, it is evident that additional measures of the glucose trajectory

(e.g., $n_{\text{obs}} = 2$ or $n_{\text{obs}} = 3$) lead to less conservative estimates, facilitating the customization of physical activity plans to individual needs.



(a) Factual glucose trajectories as a function of insulin sensitivity and β-cells mass representing prediabetic trajectories (Group A), reversible progression to T2D (Group B) and non-reversible progression to T2D (Group C) over a span of five years.



(b) Examples of factual glucose trajectories of individuals progressing to T2D over a span of five years (black lines) with their counterfactual trajectories, corresponding to the effect of all alternative physical activities reverting to prediabetes (gray lines).

Fig. 2: Glucose trajectories for individuals with $w = 70$ kg and pa of 90 min, 1/wk at 70% intensity, starting from $g_0 = 100$ mg/dl and showing prediabetic measurements at six months ($n_{\text{obs}} = 2$ with $t_1 = 0.5$ years, left panels) and at one year ($n_{\text{obs}} = 3$ with $t_1 = 0.5$ and $t_2 = 1$ year, right panels).

Indeed, two additional glucose measurements at six months and one year (right panels) reduces the number of compatible factual trajectories and the number of

potential counterfactual plans compared to a single additional measurement at six months (left panels). In turn, the extraction of more specific counterfactual physical activities avoids creating exercise overload for the individual.

4 Discussion and Conclusion

This study employed a counterfactual inference approach [12] applied to an ODEs model of blood glucose regulation [3] to examine the hypothetical impact of tailored physical activity recommendations on T2D progression in individuals at elevated risk, for example those with prediabetes. The study exploited patient-specific parametrization of the model (Table 1) to examine variability arising from differences in pancreatic β-cells functionality and insulin sensitivity. Impairments in these mechanisms, both genetic and acquired (e.g., inflammation, oxidative stress, free fatty acid levels, gluco- and lipo-toxicity), can affect insulin secretion and cells responsiveness to insulin, significantly contributing to the pathogenesis of T2D [9]. Three groups of five-year glucose trajectories were analyzed (Figure 2a). Particular attention was given to Group B, composed of individuals whose progression to T2D could have been prevented via customized physical activity plans. Within this group, the availability of longitudinal measures improved the individual characterization, thus resulting in a more precise definition of counterfactual physical activity plans (Figure 2b). Ultimately, this analysis evaluated the effectiveness of a given physical activity plan a posteriori, offering insights for more tailored suggestions in similar future subjects.

The study presents some limitations. First, only a subset of initial conditions and model parameters have been varied within a limited range to limit computational complexity. Further studies may address the variation of additional parameters and initial conditions (e.g., the initial value of insulin sensitivity, S_{I0}), as well as a finer sampling grid to extend the explored feature space. Furthermore, the proposed approach is subject to the same inherent limitations of the dynamical T2D progression model, for example, in relation to the inability of the model to differentiate between subjects of different age and sex assigned at birth or between different diets. To address these limitations, additional equations and features need to be incorporated and evaluated. Lastly, both the model and the effect of the retrieved counterfactuals should be validated with real data based on clinical trials to assess the feasibility of the approach.

Although applied here in the context of T2D progression, the proposed methodology (**Algorithm 1**) can be easily adapted to any numerically solvable ODEs model with one or more observations of the evidence over time. Noticeably, performing causal inference on a system of ODEs leveraged dynamical cause-and-effect relationships derived from prior physics-informed knowledge, rather than purely associative relationships [1].

Acknowledgments. This work was supported by the European Union (EU) and by the Swiss State Secretariat for Education, Research and Innovation (SERI) through the Project PRAESIIDIUM "Physics Informed Machine Learning-Based Prediction

and Reversion of Impaired Fasting Glucose Management" (call HORIZON-HLTH-2022-STAYHLTH-02) under Grant 101095672. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the EU and SERI. The EU and SERI can not be held responsible for them. Marta Lenatti is a PhD student enrolled in the National PhD in Artificial Intelligence, XXXVIII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Bongers, S., Blom, T., Mooij, J.M.: Causal modeling of dynamical systems (2022), arXiv:1803.08784
2. Böhm, A., Weigert, C., Staiger, H., Häring, H.U.: Exercise and diabetes: relevance and causes for response variability. Endocrine **51**(3), 390–401 (2016)
3. De Paola, P.F., Borri, A., Dabbene, F., Keshavjee, K., Palumbo, P., Paglialonga, A.: A novel mathematical model for predicting the benefits of physical activity on type 2 diabetes progression (2024), arXiv:2404.14915
4. De Paola, P.F., Borri, A., Paglialonga, A., Palumbo, P., Dabbene, F.: A model-based approach for glucose control via physical activity. In: Studies in Health Technology and Informatics (2025), in press
5. Feuerriegel, S., Frauen, D., Melnychuk, V., Schweisthal, J., Hess, K., Curth, A., Bauer, S., Kilbertus, N., Kohane, I.S., van der Schaar, M.: Causal machine learning for predicting treatment outcomes. Nature Medicine **30**, 958–968 (2024)
6. Herman, J., Usher, W.: SALib: An open-source python library for sensitivity analysis. The Journal of Open Source Software **2**(9) (2017)
7. Iwanaga, T., Usher, W., Herman, J.: Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. Socio-Environmental Systems Modelling **4**, 18155 (2022)
8. Khan, T.A., Field, D., Chen, V., Ahmad, S., Mejia, S.B., Kahleová, H., et al.: Combination of multiple low-risk lifestyle behaviors and incident type 2 diabetes: a systematic review and dose-response meta-analysis of prospective cohort studies. Diabetes Care **46**(3), 643–656 (2023)
9. Khin, P.P., Lee, J.H., Jun, H.S.: Pancreatic beta-cell dysfunction in type 2 diabetes. European Journal of Inflammation **21** (2023)
10. Multerer, L., De Paola, P.F., Lenatti, M., Paglialonga, A., Azzimonti, L.: Homogenization of ordinary differential equations for the fast prediction of diabetes progression (2024), arXiv:2412.16261
11. Pearl, J.: Causality: models, reasoning and inference. Cambridge University Press, USA, 2nd edn. (2009)
12. Pearl, J.: Probabilities of causation: three counterfactual interpretations and their identification. In: Probabilistic and Causal Inference, pp. 317–372. ACM, New York, NY, USA, 1 edn. (2022)
13. Ross, L.M., Slentz, C.A., Kraus, W.E.: Evaluating individual level responses to exercise for health outcomes in overweight or obese adults. Frontiers in Physiology **10** (2019)
14. The Diabetes Prevention Program (DPP) Research Group: The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care **25**(12), 2165–2171 (2002)

15. World Health Organization: WHO guidelines on physical activity and sedentary behaviour. World Health Organization (2020)
16. Zaffalon, M., Antonucci, A., Cabañas, R., Huber, D., Azzimonti, D.: Efficient computation of counterfactual bounds. International Journal of Approximate Reasoning **171**(C), 109111 (2024)