

## RESEARCH ARTICLE

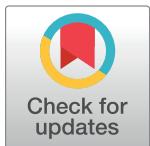
# A model for understanding the causes and consequences of walking impairments

Michael H. Schwartz<sup>1,2\*</sup>, Katherine M. Steele<sup>3</sup>, Andrew J. Ries<sup>1</sup>, Andrew G. Georgiadis<sup>1,2</sup>, Bruce A. MacWilliams<sup>4,5</sup>

**1** Center for Gait and Motion Analysis, Gillette Children's Specialty Healthcare, St. Paul, Minnesota, United States of America, **2** Department of Orthopedic Surgery, University of Minnesota, Minneapolis, Minnesota, United States of America, **3** Department of Mechanical Engineering, University of Washington, Seattle, Washington, United States of America, **4** Movement Analysis Laboratory, Shriners Hospitals for Children, Salt Lake City, Utah, United States of America, **5** Department of Orthopaedics, University of Utah, Salt Lake City, Utah, United States of America

✉ These authors contributed equally to this work.

\* [schwa021@umn.edu](mailto:schwa021@umn.edu)



## OPEN ACCESS

**Citation:** Schwartz MH, Steele KM, Ries AJ, Georgiadis AG, MacWilliams BA (2022) A model for understanding the causes and consequences of walking impairments. PLoS ONE 17(12): e0270731. <https://doi.org/10.1371/journal.pone.0270731>

**Editor:** Tomoyoshi Komiyama, Tokai University School of Medicine, JAPAN

**Received:** June 15, 2022

**Accepted:** December 4, 2022

**Published:** December 28, 2022

**Copyright:** © 2022 Schwartz et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are found in the [Supporting information](#) files (data filename = sample\_data.RData).

**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Walking is an important skill with positive impacts on health, function, and well-being. Many disorders impair walking and its positive impacts through a variety of complex and interrelated mechanisms. Any attempt to understand walking impairments, or the effects of interventions intended to treat these impairments, must respect this complexity. Therefore, our main objectives in conducting this study were to (1) propose a comprehensive model for quantifying the causes and consequences of walking impairments and (2) demonstrate the potential utility of the model for supporting clinical care and addressing basic scientific questions related to walking. To achieve these goals, we introduced a model, described by a directed acyclic graph, consisting of 10 nodes and 23 primary causal paths. We gave detailed descriptions of each node and path based on domain knowledge. We then demonstrated the model's utility using a large sample of gait data ( $N = 9504$ ) acquired as part of routine care at a regional referral center. We analyzed five relevant examples that involved many of the model's nodes and paths. We computed causal effect magnitudes as Shapley values and displayed the overall importance of variables (mean absolute Shapley value), the variation of Shapley values with respect to underlying variables, and Shapley values for individual observations (case studies). We showed that the model was plausible, captured some well-known cause-effect relationships, provided new insights into others, and generated novel hypotheses requiring further testing through simulation or experiment. To aid in transparency, reproducibility, and future enhancements we have included an extensively commented *R*markdown file and a deidentified data set.

## 1. Introduction

### 1.1 Motivation and purpose

Walking is important. It is the fundamental mode of human movement, contributes to independence, plays a role in activity, participation, and social relations, and promotes health.

Walking impairments arise from medical conditions like cerebral palsy, osteoarthritis, stroke, obesity, traumatic injuries, and many others. These impairments are treated with a variety of approaches, such as physical therapy, orthoses, oral and injected pharmaceutical agents, orthopedic surgery, and neurosurgery. To improve or correct walking impairments and their sequelae it is necessary to (1) identify the cause of the impairment, (2) have access to treatments that impact the identified cause, and (3) skillfully deliver the proper dose of the treatment.

Consider a patient with a severely osteoarthritic hip. The patient has a slow, unstable, Trendelenburg gait, has trouble conducting activities of daily living, has stopped participating in leisure activities, and reports that these problems lower their independence and happiness. The primary causal path from the osteoarthritic hip to the impairments seems clear (cartilage degradation), a well-established and effective treatment is available (total hip arthroplasty), and the dose is not in question (one hip). As a result, a total hip arthroplasty for this patient is likely to result in large predictable improvements to walking, mobility, and quality-of-life [1]. Of course, this is a simplification. The patient may have developed weakness or contractures while walking with compensations for many years, may have risk factors for a poor outcome from surgery, and may have underlying causes of the osteoarthritis that are not addressed by joint replacement. This reinforces the idea that movement impairing conditions are inherently complex.

Now consider a patient with cerebral palsy (CP) who has spasticity, poor motor control, weakness, muscle contractures, bony torsions, and unstable feet. The patient presents with comparable gait, mobility, and quality-of-life complaints as the osteoarthritic patient. Which of the patient's underlying orthopedic and neurological problems are the cause of their impairments? What treatments are available for these underlying causes? What dose of correction is appropriate? Unfortunately, treatment of impairments arising from CP result in modest and unpredictable outcomes that have been stagnant for at least the last several decades [2]. We believe that the difficulty in identifying causes of gait impairments is an important contributor to this unsatisfactory situation.

In this study we will propose a comprehensive model for quantifying causes and consequences of walking impairments and demonstrate the potential utility of the model for supporting clinical care decisions and addressing scientific questions.

## 1.2 Causal inference

Causal inference has been carefully and succinctly defined many times, including for the International Encyclopedia of the Social and Behavioral Sciences (Second Edition). Quoting extensively below [with adaptations for our application to walking impairments in brackets]:

*“Causal inference refers to an intellectual discipline that considers the assumptions, study designs, and estimation strategies that allow researchers to draw causal conclusions based on data. . . The dominant perspective on causal inference in statistics has philosophical underpinnings that rely on consideration of counterfactual states. In particular, it considers the outcomes that could manifest given exposure to each of a set of [causal factors]. Causal effects are defined as comparisons between these ‘potential outcomes.’ For instance, the causal effect of [femoral anteversion] on [in-toeing] would be defined as a comparison of the [in-toeing] that would be measured [with one level of femoral anteversion] with the [in-toeing] that would be measured [with a different magnitude level of femoral anteversion]. The challenge for causal inference is that we are not generally able to observe both of these states: at the point in time when we are measuring the [in-toeing], each individual . . . has [only one level of femoral anteversion]”*

[3].

We emphasize that the key difficulty for causal inference is that we can never simultaneously observe the actual and counterfactual states (e.g., the same limb with two different levels of anteversion). As a result, causal conclusions are not falsifiable. Instead, their validity is supported or refuted by indirect evidence of a logical or mechanistic nature. Despite this fundamental challenge, causal inference has seen an explosion in use over the past two decades, thanks in large part to the clarification of its theoretical underpinnings, extensive testing and simulation, and the proliferation of powerful analytical tools. A search of <https://app.dimensions.ai> for the term “*causal inference*” showed that ~190,000 out of 200,000 all-time hits (~93%) were published after the year 1999 [4]. It is clear that causal inference techniques are an exceptionally useful and important modern analytical tool.

### 1.3 Causal identification

We will not recapitulate the entire field of causal inference here, since there are countless outstanding resources written by experts in the field. For a relatively non-technical yet rigorous and highly entertaining overview, we direct the reader to *The Book of Why* (Pearl and MacKenzie, 2018). The process of causal identification starts with proposing an underlying assumed model. One way such a model can be described is in terms of a directed acyclic graph (DAG). A DAG consists of a set of nodes, representing variables, connected to one or more other nodes by arrows. These arrows represent direct causes of one variable on another. The simple diagram  $A \rightarrow B$  means: “*A causes B*” (e.g., the wind causes the leaves to flutter, not the other way around, as the 6-year-old version of the first author believed). Representing a set of causal hypotheses as a DAG has several strengths, including clear visual communication and the ability to rigorously define requirements for the estimation of causal effects. The latter is largely the result of the work of Pearl, and is the key tool in this study for examining causal effects related to walking impairments [5]. A DAG also allows the identification of conditional independencies—partial correlations that must be zero (or small, in practice). These independencies can be tested to examine the plausibility of the proposed DAG. Satisfying the conditional independencies is a necessary, but not sufficient condition to show accuracy of the proposed model.

**Total and direct effects.** We will be examining both direct and total effects in this study. The *direct effect* of exposure A on outcome B is the effect passing through the causal arrow  $A \rightarrow B$ . However, there may also be indirect effects of A on B. For example, perhaps the diagram consists of the paths  $A \rightarrow B$  and  $A \rightarrow M \rightarrow B$ . In this case, A has a direct effect on B and an indirect effect on B, mediated by M. The *total effect* of A on B accounts for causal effects flowing through all paths—both direct and mediated.

**Adjustment sets.** Given a causal diagram we need a way to isolate the causal effect of a variable A (sometimes called an “*exposure*” or a “*treatment*”) on variable B (sometimes called an “*outcome*”). This is done by finding a set of variables called an adjustment set. The effect of A on B is causal only after conditioning for the adjustment set, which consists of the other variables that influence the effect of A on B. In an experimental setting, this can be achieved by controlling for the adjustment set variables through experimental design. In an observational setting, we cannot control events. Instead, we include the adjustment set variables in a model. There are mathematically rigorous rules that allow us to find adjustment sets or determine that no such set exists [5]. The adjustment set required to determine total effects may be different from the one required to determine direct effects. A proper adjustment set needs to be determined for each variable in order to avoid misrepresenting a non-causal effect for a causal one. This misrepresentation is called the “*Table 2 Fallacy*” [6].

### 1.4 Causal estimation

We use a computational model to turn the causal arrows in the DAG into numerical values. To do this we first represent the variables, which are abstract concepts like mobility, as numbers (e.g., 1.35, 2.7, 42, ...) or labels (e.g., mild, moderate, severe, ...). The concept of the variable and its realization as a number or label will be considered exchangeable. If there is a well-understood mechanism underlying the causal path, the computational model can be explicit. For example, suppose a causal diagram has a direct arrow from  $A \rightarrow B$ , and we know the effect of  $A$  on  $B$  is linear. Then we can model the path as  $B = \beta * A$ . In this case, the causal effect is simply the coefficient  $\beta$ . If only the world was so simple.

The relationships we are examining in the model we are proposing do not generally have well understood mechanistic relationships that can be expressed as algebraic formulae. Also, the number of relationships as well as their complexity and unknown multi-way interactions presents a challenge. So, rather than proposing explicit relationships for each causal path, we take the approach of using a flexible computational engine that can accurately model these relationships. We choose to use Bayesian Additive Regression Trees (BART) as the computational engine [7]. Using BART for causal inference has been shown to produce accurate and precise estimates of causal effects across a wide range of real and synthetic datasets [8, 9]. The reason for this exceptional performance is a result of BART's ability to flexibly recreate the shape of the outcome's response surface.

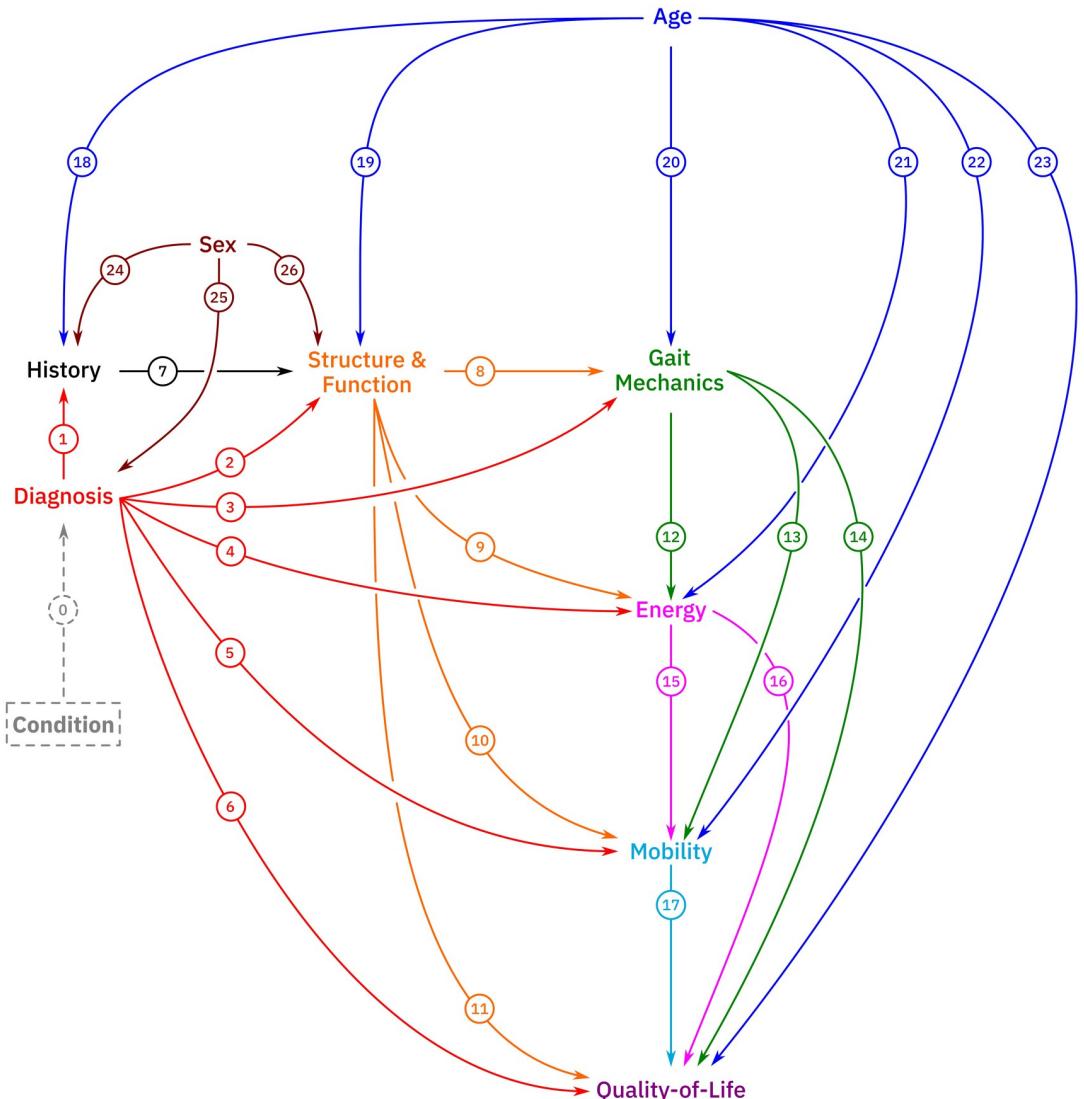
**Interpreting results.** In a linear regression it is straightforward to interpret an effect—it is the coefficient of the variable in the regression equation ( $\beta$  in the above example). A disadvantage of BART is that, as a “*black box*” machine learning algorithm, extracting interpretable causal effect magnitudes requires additional analysis. There are many proposed solutions to this problem, including partial dependence plots and accumulated local effects. In this study we have chosen to use Shapley values to represent the causal effect magnitudes [10]. A variable's Shapley value is the amount that it contributes to the model prediction for the outcome. Shapley values are frequently computed relative to the sample mean for the outcome—in which case the sum of the sample mean outcome plus all Shapley values for all variables equals the predicted outcome. Shapley values are not necessarily constant, but can depend on other covariates. As an analogy, we return to the linear regression example above  $A \rightarrow B$  where  $B = \beta * A$ . In this example  $\beta$  is the causal effect and has a constant value. But what if the value of  $\beta$  depended on the value of  $A$ ? In that case we would have  $B = \beta(A) * A$ . In fact, in a more complex situation,  $\beta$  could depend on several variables, not just  $A$ . In what follows, we present Shapley values three ways. If  $S_{age}$  is the Shapley value for the variable age, we compute and display the mean absolute Shapley value over all observations (importance plot). We also examine how age affects the Shapley value by displaying  $S_{age}(age)$  (dependence plot). Lastly, we compute Shapley values for important variables on an individual basis (case studies).

In the following sections we will introduce a model for understanding the causes and consequences of walking impairments. We will describe the model in detail and present five examples where the model is implemented using data collected during routine clinical gait analysis assessments.

## 2. Materials and methods

### 2.1 The model

The proposed causal model consists of 10 nodes and 23 primary causal paths (Fig 1). For clarity, the depiction of the model omits some details that are described below and may be found



**Fig 1. The proposed causal model in schematic form.** The proposed model nodes include *Condition (latent)*, *Diagnosis*, *History*, *Age*, *Sex*, *Structure and Function*, *Gait Mechanics*, *Mobility*, *Energy*, and *Quality-of-Life*. Extensive details of proposed nodes are found in the text. Descriptions of the paths are found in [S1 Appendix](#). A full implementation of the model (*Markdown* file) and a de-identified dataset can be found in the [S1 File](#).

<https://doi.org/10.1371/journal.pone.0270731.g001>

in an included *R* *Markdown* file (n.b., we will use an italicized, monospaced font with a grey background to indicate software). The model captures causal relationships at a snapshot in time. In other words, all nodes, including *History* (n.b., nodes will be capitalized and displayed in italics throughout), are evaluated at  $time = t$ . For example, a patient may present with a hamstrings contracture and walk in a crouch gait pattern. The model allows us to examine the effect of hamstrings contracture on crouch gait. However, it is possible that continuing to walk in such a pattern could result in worsening of hamstring contracture, or perhaps the patient might undergo a hamstring lengthening surgery to improve the contracture. These changes would be reflected in a model at  $time = t + 1$ .

## 2.2 Model nodes

*Condition* is a latent variable capturing the exact details of the patient's medical condition. Note that *Condition* is a placeholder term for anything causing a change in body structure or function. This could include injuries, diseases, genetic mutations, malnutrition, aging, etc.

*Diagnosis* is a surrogate for the *Condition* variable. We do not expect *Diagnosis* alone to capture the full details of the patient's condition. However, this coarse parameter allows us to adjust for some of the missing condition-dependent variables we do not measure.

*History* includes events that can significantly influence body structures and function, such as prior treatments. Since the model uses the exam date as a starting point, we do not model the causes of the prior treatments. Therefore, within the *History* node, we connect prior treatments to a latent variable of unknown common causes ( $Prior_j \leftarrow Unknown \rightarrow Prior_k$  for all  $j, k$ ). Note that this detail is not shown on the schematic version of the causal model but can be found in the included *Rmarkdown* file.

*Structure and Function* is shorthand for Body Structures and Functions as described in The International Classification of Functioning, Disability and Health (ICF) framework [11]. We are specifically thinking of variables like spasticity (ICF category b750 –Motor reflex functions and b735 –Muscle tone functions), motor control (b760 –Control of voluntary movement functions), strength (b730 –Muscle power), range-of-motion, bony alignment, muscle contracture, and joint health (s740 –Structure of pelvic region, s750 –Structure of lower extremity, s760 –Structure of the trunk, and s770 –Additional musculoskeletal structures related to movement). The *Structure and Function* node is the core of the proposed causal model since elements of this node are the direct targets of treatments, and treatment effects are largest at this level [12].

*Gait Mechanics* are the elements of dynamics pertaining to the motion of segments and joints during gait (usually walking, but running, skipping, and other gaits could be considered). These data can come from three-dimensional gait analysis, video, or observational analysis, depending on the questions being investigated. As long as the concept of interest can be coded as a number or category, it can be included in the model. *Gait Mechanics*, as we mean it, appears in the ICF as both a body function (b770 –Gait pattern functions) and as an activity (d450 –Walking). When choosing variables to represent *Gait Mechanics* additional choices need to be made about the relationships among those variables. Variables in this node are related to one another through definition as well as the geometric and physical constraints of the lower extremity and walking. Given the topology of the assumed causal model, it can be shown that these relationships (e.g., *Structure and Function* → *Gait Mechanics* or *Gait Mechanics* → *Energy*) do not impact important total causal effect sizes. However, they do impact implied conditional independencies. For example, if minimum hip flexion and hip flexion range of motion during gait (related algebraically) are included as part of the *Gait Mechanics* node, then a shared cause must be added (i.e., minimum hip flexion ← *Unknown* → range-of-motion hip flexion). We therefore add an unknown factor to this node that is connected to all *Gait Mechanics* variables ( $Gait\ Mechanics_j \leftarrow Unknown \rightarrow Gait\ Mechanics_k$  for all  $j, k$ ). Note that this detail is not shown on the schematic version of the causal model (Fig 1) but can be found in the included *Rmarkdown* file.

Cadence, step-length, and speed do not appear explicitly in the causal model. Instead, they are considered part of *Gait Mechanics*, but merit special attention. Speed is the product of cadence and step-length. These are part of ICF Walking (d450 –Walking), which is defined as “*Moving along a surface on foot, step by step, so that one foot is always on the ground, such as when strolling, sauntering, walking forwards, backwards, or sideways.*” Alert readers will note that the ICF defines walking, in part, as “*walking*”. The trio of cadence, step-length, and speed

are not independent. Any two can be chosen freely and fully determine the third. We choose cadence and step-length as the causes of speed since, to walk (1) we turn muscles on and off in a sequence and intensity that generates joint movements, (2) the joint movements result in a step of a given length, and (3) we repeat the step-length generating activation pattern at a given cadence. Thus, we choose to treat cadence and step-length as the independent causal variables. Including speed is redundant. However, there are meaningful effects that are more sensibly expressed as speed effects rather than as combined cadence and step-length effects. For example, the inability to keep up with peers while walking in the community, a frequent patient complaint, is most clearly expressed as a speed problem. Other effects are more naturally expressed in terms of cadence and step-length, such as the effect these variables have on energy [13]. Thus, we include cadence, step-length, and speed in the causal model and include the condition that  $\text{cadence} \rightarrow \text{speed} \leftarrow \text{step-length}$ . It is important to keep this redundancy in mind when working with the model, since attempts to independently assign speed, cadence, and step-length will result in computational and interpretation challenges. This also highlights the fact that the model schematic does not contain every detail required for computation. These details can be found in the included *R* *Markdown* file.

*Energy* is the metabolic energy required for walking (ICF b5409 –General metabolic functions unspecified or b789 –Movement functions other specified and unspecified). We recommend against using energy per unit distance—often called “*energy cost*”, which is obtained by dividing the time-rate of energy consumption (power) by walking speed. Problems arise at slow speeds, where energy cost is dominated by the  $\text{speed}^{-1}$  term, and the impact of the metabolic term becomes trivial.

*Mobility* is a broad term referring to a range of functions and abilities related to moving around within the home or community (ICF d410-429 –Changing and maintaining body position, d430-449 –Carrying, moving, and handling objects, d450-469 –Walking and moving, and d47-489 –Moving around using transportation). The elements of mobility that are important are patient-specific, and influenced by culture, environment, and many other factors.

*Quality-of-Life* may be clear conceptually but is a vexingly difficult quantity to define and measure. For one patient, walking faster may have a large impact on their quality-of-life. For another, walking speed may not matter, but gait appearance does. Despite the central importance of this measure, it is the most likely to be influenced by factors not represented in the diagram, such as social relations, socio-economic status, psychiatric disturbances, job status, and so on. There are some successful and widely used measures for this domain, such as the World Health Organization Quality of Life Assessment and the Diener Satisfaction with Life Scale [14, 15]. We leave *Quality-of-Life* in the diagram as a reminder of its central importance, while simultaneously acknowledging the limitations of the model regarding this node.

### 2.3 Causal paths

Each path in the model indicates the possible existence of a causal effect. It is only necessary to find one possible example of the path for it to be included in the model. The magnitude of the effect may turn out to be trivially small, but that is for the data to show. Given this situation, one might ask “*why not allow possible causal paths between all pairs of variables?*”. The answer is that such a fully connected model produces no implied conditional independencies, and therefore, cannot be falsified. It also ignores the nature of the true causal relationships among variables, which may result in the introduction of nonexistent and biasing paths. We do not indicate if an effect is linear, quadratic, or some other form. Nor do we describe the existence or nature of interactions. Those determinations are part of the causal estimation described in

section 1.4, Causal Estimation. Brief examples of possibilities for each of the 23 causal paths in the proposed model can be found in [S1 Appendix](#).

## 2.4 Missing nodes and paths

No causal model can include every node and path. Some missing nodes that seem important include socio-economic status, cognition, family dynamics, internal motivation, pain, or other medical conditions. Missing nodes and paths have two possible impacts. First, they may increase unexplained variance in an outcome measure. Second, and generally more importantly, they may bias the causal estimates. It is difficult to predict whether the missing nodes and paths in a model are biasing or not. There are methods to estimate how much bias missing paths would need to introduce to change causal conclusions [16]. We will not perform these sorts of analyses in this study.

## 2.5 Model realization

**Data.** This study was reviewed and authorized by the University of Minnesota institutional review board review (STUDY00012420). All experiments were performed in accordance with relevant guidelines and regulations. Informed consent for use of medical records was obtained at the time of service from all participants or their legal guardians. An option to rescind this permission is offered to patients at every visit to our center.

We used data extracted from our laboratory's historical clinical database collected between 1994 and 2022. Data consisted of measures collected for clinical standard-of-care evaluations.

For the computational models, missing continuous data was left as missing. Missing categorical data was neither omitted nor imputed, but instead assigned its own category ("Missing"). For causal effects dependency plots, we did not plot the "Missing" category since we are not modeling the pattern of missingness, and hence this category has no clear causal meaning. Handling of missing data in future applications of the model (e.g., hypothesis testing) must be carefully considered. For example, in our data we know that missing strength values are associated with more significant physical and cognitive impairments due to a patient's inability to understand or cooperate with the examination.

**Variables.** To turn the conceptual causal model into a practical tool, we needed to realize the nodes as measured values. The variables contained in the model nodes were chosen to be comprehensive and commonly measured in clinical gait analysis centers to promote replication efforts and model extensions.

*Diagnosis.* Assigned diagnosis lumped into the categories cerebral palsy, myelomeningocele, developmental variants, traumatic brain injury, and other (see [S3 Appendix](#) for listing of "other" diagnoses).

*History.* Prior treatments lumped into the categories selective dorsal rhizotomy, neurolysis (e.g., botulinum toxin type A or phenol injection), rectus femoris transfer, femoral derotation osteotomy, tibial derotation osteotomy, distal femoral extension osteotomy, foot and ankle bone surgery, psoas lengthening, hamstrings lengthening, adductor lengthening, calf muscle lengthening, patellar advancement, foot and ankle soft-tissue surgery. These 13 account for around 90% of the treatments performed on patients in our clinical database. The causal effects of these treatments have been evaluated previously [17].

*Sex.* Sex as indicated by the patient's medical record.

*Age.* Age ex-vivo. Note that when dealing with very young individuals or incipient walkers, it might be more appropriate to consider an adjusted age that accounts for premature birth.

*Structure and function.*

- Maximum range-of-motion: hip flexion, hip extension, hip abduction (knee extended), popliteal angle (unilateral), knee flexion, knee extension, extensor lag, ankle dorsiflexion (knee extended), ankle plantarflexion. All measured with handheld goniometry.
- Long bone torsion: trochanteric prominence angle test, bimalleolar axis angle. All measured with manual goniometry.
- Weightbearing foot alignment: forefoot ab/adduction, forefoot in/eversion, hindfoot var/valgus, midfoot position (subjectively assessed). Raw data also included severity scores for these measures, which have been dropped for this analysis.
- Spasticity: overall lower extremity spasticity defined as first component score from polychoric principal component analysis applied to spasticity of the adductors, hamstrings, hip flexors, plantarflexors, posterior tibialis, and rectus femoris [18]. The resulting measure is a z-score with respect to the sample.
- Strength: overall lower extremity strength defined as first component score from polychoric principal component analysis applied to manual muscle strength testing of abdominal, back extensors, hip abductors, hip adductors, hip extensors, hip flexors, knee extensors, knee flexors, ankle plantarflexors, anterior tibialis, posterior tibialis [19]. The resulting measure is a z-score with respect to the sample.
- Static selective motor control: overall lower extremity static selective motor control defined as first component score from polychoric principal component analysis applied to subjectively assessed measures of isolated static motor control (0-fully patterned, 1-partially patterned, 2-normal) of abdominal, back extensors, hip abductors, hip adductors, hip extensors, hip flexors, knee extensors, knee flexors, ankle plantarflexors, anterior tibialis, posterior tibialis. The resulting measure is a z-score with respect to the sample.
- Dynamic motor control: The Dynamic Motor Control Index during Walking [20]. This measure requires electromyography data. For many years we did not routinely collect electromyography data during post-treatment visits, resulting in a greater amount of missing data for this measure.

*Gait mechanics.* Kinematic parameters derived from three-dimensional gait analysis: mean pelvic tilt, pelvic tilt range-of-motion (ROM), minimum stance-phase hip flexion, mean stance-phase hip rotation, minimum stance phase knee flexion, knee flexion at initial contact, swing phase knee flexion ROM, maximum swing phase knee flexion, mean stance knee rotation, ankle dorsiflexion at initial contact, mean stance ankle dorsiflexion, mean swing ankle dorsiflexion, mean stance foot progression, dimensionless step-length, dimensionless cadence, dimensionless speed. All kinematics are obtained from the Vicon plug-in gait model with functional hip and knee axis [21, 22]. For data collected prior to 2006, we did not compute a functional knee axis and therefore do not report a knee rotation value. Speed, cadence, and step-length nondimensionalized following Hof [23]. Data represent the average of at least three (and generally fewer than 7) trials of over-ground walking at self-selected speed. This reflects our standard clinical protocol.

*Energy.* Net energy consumption during walking at self-selected speed. The value is nondimensionalized and expressed as a z-score with respect to speed-matched typically-developing controls [24]. For simplicity we will refer to this as net energy consumption.

*Mobility.* Functional Assessment Questionnaire Transform (FAQt) [25]. Demonstrated mobility measures like the GMFM-66 would be preferable, but these data are not available for

most of the patients in this study. The FAQt has been shown to be strongly correlated with the GMFM-66 ( $r \sim .60$ ).

**Quality-of-Life.** Domain scores from the GOAL survey: Activities of Daily Living and Independence, Body Image and Self-Esteem [26].

**Software.** We use the *dagitty* package in *R* to build the DAG [27]. Programming details are in the included *Rmarkdown* file (electronic addendum).

**Deriving adjustment sets.** We derived adjustment sets using the provided functions in *dagitty*.

**Computational engine.** We used BART models to compute the causal effects. For outcome variables expressed as continuous or dichotomous variables we used the *bartMachine* package [28]. We enabled the use of missing data, which BART handles without imputation [29]. Note that outcomes expressed as polytomous variables can be predicted using the *BART* package [30]. We used default settings for all BART models, which have been shown to be nearly optimal across a wide range of conditions. We found only  $\sim 1\text{--}3\%$  improvement in out-of-sample performance from cross-validation optimized models.

**Quantifying causal effect magnitude with Shapley values.** Shapley values were computed using the *fastShap* package [31]. We present Shapley values in three different forms: (1) importance plots display mean absolute Shapley values and give an overall picture of the causal importance of a variable, (2) dependence plots show how the Shapley value varies over the domain of the causal variable and (3) case studies show Shapley values of chosen variables for individual patients.

## 2.6 Plausibility of the model

Model plausibility was tested using implied conditional independencies. In the case of the proposed model, most of these involved large numbers of conditioning variables (2 tests with 0 conditioning variables, 48 with 1, 165 with 15, 280 with 20, and 1 with 38). The calculations are unreliable in this situation, so we only evaluated conditions with 0 or 1 conditioning variables [32]. Furthermore, many of the tests did not provide meaningful plausibility information since they were essentially tests of the relationship between speed, cadence, and step length, which is fulfilled by definition. To compute test values we first generated a polychoric correlation matrix for the variables using the *lavaan* package [33]. Note that this step was necessary since the implied conditional independencies involved a mix of continuous and ordered data. The diagnosis variable was further lumped into two categories—CP ( $\sim 80\%$  of observations) and Other. This dichotomization was necessary since diagnosis is an unordered category.

## 2.7 Examples

We demonstrate the utility of the proposed causal model with five examples. The purpose of these examples is to show how to use the model, highlight the model's potential clinical and scientific value, and provide some face validity. There are many different combinations of exposures and outcomes we could consider—each producing dozens of graphs and tables. Rather than giving an encyclopedic treatment here we are providing an *Rmarkdown* file and a demonstration data set of deidentified observations so that readers can run the *Rmarkdown* file to create facsimiles of the plots presented here and can modify the code to explore, enhance, test, or utilize the causal model. For each example we (1) chose an exposure node and outcome measure, (2) determined the proper adjustment set for valid causal inference, (3) built a BART model for the outcome based on the adjustment set, and (4) examined a combination of importance plots, dependence plots, and case studies.

**Example 1: Total effects of *Structure and Function* on mean stance foot progression.**

The effect of impairments at the *Structure and Function* level on impairments at the *Gait Mechanics* level is the primary problem of interest in most clinical gait studies, and directly drives treatment decisions. The broader goals of treatment may include improving *Energy*, *Mobility*, or *Quality-of-Life*, but the treatment assignment process is largely based on assumptions about how structural impairments cause gait impairments. These assumptions are *ad hoc*, rarely based on evidence, and essentially never based on causal evidence, since the gait literature contains (almost) none. To demonstrate the application of the proposed model to this causal path we examined the total effects of *Structure and Function* on in- and out-toeing, which is a common problem of concern to patients and is frequently treated with surgery. We quantified in- and out-toeing by mean stance foot progression. The adjustment set consists of variables belonging to the *Structure and Function*, *Age*, and *Diagnosis* nodes [S2 Appendix].

**Example 2: Total effects of *Structure and Function* on FAQt.** *Mobility* is important to a patient and is often discussed in relation to treatment decisions. We have chosen to quantify *Mobility* using the FAQt. Note that some components of the FAQt are clearly mediated by *Gait Mechanics* (e.g., overall walking ability and climbing stairs) and some are not (e.g., kicking, hopping, and jumping). The adjustment set consists of variables belonging to *Structure and Function*, *Age*, and *Diagnosis* nodes. Note that it would be wrong to adjust for *Gait Mechanics*, which is a mediator (i.e., there is a path consisting of *Structure and Function* → *Gait Mechanics* → *Mobility*) [S2 Appendix].

**Example 3: Total effect of *Gait Mechanics* on net energy consumption.** The net energy consumption during walking for children with cerebral palsy is high—averaging over twice that of typically developing controls. We quantify *Energy* walking -resting energy consumption (time rate of energy) nondimensionalized and expressed as a z-score with respect to speed matched controls [24]. Common treatments show limited effectiveness to reduce net energy consumption. In a previous study, overall kinematic deviations were shown to contribute substantially to elevated net energy consumption [34]. By examining the effects of *Gait Mechanics* on net energy consumption we can determine which specific kinematic deviations are the most impactful. The adjustment set consists of variables belonging to the *Gait Mechanics*, *Structure and Function*, *Age*, and *Diagnosis* nodes [S2 Appendix].

**Example 4: Direct effect of *Age* on net energy consumption.** A particular challenge with observational data is disentangling direct effects of age from effects due to other factors, including treatment. In this example, we examine the direct effect of age on net energy consumption. By examining the direct effect of age on net energy consumption we can quantify the effect that maturation has, independent of associated changes in *Structure & Function*, *History*, or *Gait Mechanics*. Unlike the previous examples, where we estimated total effects, here we examine the direct effect of *Age*. This is an effect that is impossible to observe—hence an ideal situation for causal inference methods. The direct effect isolates the developmental role of *Age* from the concomitant impact it has on factors like gait mechanics and treatment decisions, which result in indirect causal effects. The adjustment set consists of variables belonging to the *Diagnosis*, *Structure and Function*, and *Gait Mechanics* nodes [S2 Appendix]. For this example, we analyzed only individuals diagnosed with cerebral palsy.

**Example 5: Total effects of (1) *Structure and Function* on Activities of Daily Living and Independence and (2) *Gait Mechanics* on Body Image and Self-Esteem.** Independence and self-esteem are important aspects of *Quality-of-Life*. The GOAL survey provides validated self-report measures from these domains. We use the GOAL's Activities of Daily Living and Independence domain score as a measure of independence, and the Body Image and Self-Esteem score as a measure of self-esteem. We estimate the total effects of *Structure and Function* on the Activities of Daily Living and Independence score and *Gait Mechanics* on the Body Image

and Self Esteem score. The adjustment set for the total effects of *Structure and Function* on Activities of Daily Living and Independence consists of variables belonging to the *Age*, *Diagnosis*, and *Structure and Function* nodes. The adjustment set for the total effect of *Gait Mechanics* on Body Image and Self-Esteem consists of variables belonging to the *Age*, *Diagnosis*, *Structure and Function*, and *Gait Mechanics* nodes [S2 Appendix].

For examples 1 through 3 we computed Shapley values for a random sample of 1500 observations, which is around 1/6<sup>th</sup> of the data and gives a representative picture of the response without making the plots overly dense. For example 4 we also included all observations younger than 4 years and older than 15 years in order to have sufficient density across ages. For example 5 we selected all available data.

### 3. Results

#### 3.1 Data

The database query resulted in 9504 individuals (Tables 1–5). We selected the affected side (unilateral diagnosis) or the left side (bilateral diagnosis) from each individual for further analysis. We have not seen a meaningful impact from the use of bilateral data in this model or previous causal treatment effect models, though there may be situations where using bilateral data is appropriate.

#### 3.2 Implied conditional independencies

As noted in the methods, due to the number of variables in the model and the structure of the proposed DAG, many implied conditional independencies could not be tested and those that could were largely fulfilled by definition. The conditions that were tested exhibited partial correlations less than 0.20.

#### 3.3 Examples

For each example, observations where the outcome was missing were excluded resulting in a reduced sample size in examples 1–5 (N = 9504, 5498, 7672, 7672, and 835/839, respectively).

##### Example 1: Total effects of *Structure and Function* on mean stance foot progression.

As expected, tibial torsion (estimated by bimalleolar axis angle) and femoral anteversion (estimated by trochanteric prominence test angle) were the two largest causes of mean stance foot progression (Fig 2). The next three largest causes of meaningful magnitude were related to foot deformity.

**Table 1. Age, sex, and diagnosis.**

Variable	N = 9,504
Age, Median (IQR)	10.5 (7.7–13.8)
Diagnosis, n (%)	
<i>Cerebral Palsy</i>	6,667 (70)
<i>Developmental Variant</i>	597 (6.3)
<i>Myelomeningocele</i>	314 (3.3)
<i>Traumatic Brain Injury</i>	207 (2.2)
<i>Other</i>	1,719 (18)
Sex, n (%)	
<i>Female</i>	4,247 (45)
<i>Male</i>	5,257 (55)

<https://doi.org/10.1371/journal.pone.0270731.t001>

**Table 2. History of prior treatment.**

Treatment	N
Neurolysis, n (%)	3,850 (41)
Calf Muscle Lengthening, n (%)	2,446 (26)
Femoral Derotation Osteotomy, n (%)	2,359 (25)
Selective Dorsal Rhizotomy, n (%)	2,242 (24)
Foot Bone, n (%)	1,716 (18)
Foot Soft Tissue, n (%)	1,309 (14)
Tibial Derotation Osteotomy, n (%)	1,220 (13)
Hamstrings Lengthening, n (%)	1,164 (12)
Adductor Lengthening, n (%)	891 (9.4)
Rectus Transfer, n (%)	814 (8.6)
Psoas Lengthening, n (%)	782 (8.2)
Patellar Advancement, n (%)	342 (3.6)
Distal Femoral Extension Osteotomy, n (%)	162 (1.7)

<https://doi.org/10.1371/journal.pone.0270731.t002>

The dependence of mean stance foot progression on long bone torsion appears to be approximately linear, with slopes of around 0.6 and 0.3 degrees of mean stance foot progression per degree of torsion for the tibia and femur, respectively (Fig 3). We also see the expected relationship at the forefoot, with an abducted forefoot causing external foot progression and an adducted forefoot causing internal foot progression. Identifying these foot-related effects is reassuring since they are sensible, but also pleasantly surprising, since these measures tend to be quite coarse and noisy.

As an aside, recall that Shapley values are expressed with respect to the mean over the sample, so dependence plots cross the zero line at the sample mean of the corresponding exposure value. In other words, visual inspection of Fig 3 shows that the sample mean bimalleolar axis angle is ~15°, trochanteric prominence test angle is ~35° and forefoot abduction/adduction is neutral.

Four examples from individuals—two with significant in-toeing (top row) and two with significant out-toeing (bottom row)—show different *Structure and Function* profiles leading to similar in- and out-toeing (Fig 4). This has obvious clinical importance since different treatments would be considered each of these cases.

**Example 2: Total effects of *Structure and Function* on FAQt.** Strength and motor control (static and dynamic) were the most important causal contributors to mobility as measured by the FAQt (Fig 5).

The dependence plots suggest that the difference between poor overall strength (-1 is 1 SD weaker than sample average) and good strength (+1 SD stronger than sample average) is 20 points, or 2 SD on the FAQt scale (Fig 6). While this is a large effect, the likelihood of improving overall strength by 2 SD through any intervention is low. The motor control effects are nearly as large but improving motor control is also an extremely challenging therapeutic goal.

All highly mobile patients resemble one another. They are strong and have good motor control. Those with poor mobility also fit a consistent prototype of poor strength and poor motor control. Individual examples of both high and low FAQt scores typify this finding (Fig 7).

**Example 3: Total effect of *Gait Mechanics* on net energy consumption.** The largest causal contributor to net energy consumption was initial contact knee flexion, which was twice as impactful as the next closest cause (Fig 8). Elements of stance phase knee flexion and ankle dorsiflexion were the next most important causes—equal in magnitude to one another.

**Table 3. Structure and function.**

Variable	N	Value
Maximum Hip Flexion, Median (IQR)	9,503	135 (124–135)
Maximum Hip Extension, Median (IQR)	7,422	0 (-10–5)
Maximum Hip Abduction (Knee Extended), Median (IQR)	9,499	31 (25–40)
Popliteal Angle (Unilateral), Median (IQR)	9,490	50 (40–60)
Maximum Knee Flexion, Median (IQR)	9,491	135 (135–135)
Maximum Knee Extension, Median (IQR)	9,502	0 (-2–0)
Extensor Lag, Median (IQR)	7,281	0 (0–5)
Maximum Ankle Dorsiflexion (Knee Extended, Median (IQR)	9,145	5 (0–10)
Maximum Ankle Plantarflexion, Median (IQR)	9,464	50 (42–50)
Trochanteric Prominence Test Angle, Median (IQR)	9,457	30 (20–45)
Bimalleolar Axis Angle, Median (IQR)	9,031	15 (7–20)
Forefoot Varus/Valgus, n (%)	9,504	
<i>Valgus</i>		1,260 (13)
<i>Neutral</i>		6,084 (64)
<i>Varus</i>		312 (3.3)
<i>Missing</i>		1,848 (19)
Forefoot Abd/Add, n (%)	9,504	
<i>Abducted</i>		3,391 (36)
<i>Neutral</i>		4,395 (46)
<i>Adducted</i>		1,455 (15)
<i>Missing</i>		263 (2.8)
Hindfoot Var/Val, n (%)	9,504	
<i>Valgus</i>		5,271 (55)
<i>Neutral</i>		2,983 (31)
<i>Varus</i>		572 (6.0)
<i>Missing</i>		678 (7.1)
Midfoot Cavus/Planus, n (%)	9,504	
<i>Cavus</i>		555 (5.8)
<i>Neutral</i>		2,919 (31)
<i>Planus</i>		3,875 (41)
<i>Missing</i>		2,155 (23)
Lower Extremity Spasticity, Median (IQR)	9,348	-0.48 (-0.76–0.49)
Lower Extremity Strength, Median (IQR)	8,634	0.05 (-0.69–0.76)
Lower Extremity Static Motor Control, Median (IQR)	8,660	0.31 (-0.89–1.00)
Dynamic Motor Control, Median (IQR)	4,347	82 (75–90)

<https://doi.org/10.1371/journal.pone.0270731.t003>

Shapley value dependence plots showed that initiating ground contact with the knee in a flexed position has a substantial net energy consumption penalty—around 1.5 SD difference between contacting with 10° vs. 50° of flexion (Fig 9).

It is interesting to note that walking in crouch (minimum stance knee flexion > 25°) seems to provide an energetic benefit. Previous studies, both experimental and analytical, have reported mixed and inconclusive findings regarding the relationship between energy and crouch. It is possible that this is because they failed to control or adjust for the proper confounders [35, 36]. Walking in a plantarflexed position also contributes to elevated net energy consumption. The apparent threshold effect for plantarflexion may be a consequence of using a tree-based computational model, or it may reflect a true mechanism related to the way that a heel strike promotes an energetically efficient step-to-step transition [13]. The disruption of

**Table 4. Gait mechanics.**

Variable	Value
Mean Pelvic Tilt, Median (IQR)	17 (12–22)
Range-of-Motion Pelvic Tilt, Median (IQR)	6.4 (4.5–8.9)
Minimum Stance Hip Flexion, Median (IQR)	2 (-4–10)
Mean Stance Hip Rotation, Median (IQR)	4 (-5–12)
Minimum Stance Knee Flexion, Median (IQR)	9 (3–18)
Initial Contact Knee Flexion, Median (IQR)	22 (12–32)
Range-of-Motion Knee Flexion, Median (IQR)	40 (28–51)
Maximum Swing Knee Flexion, Median (IQR)	61 (55–67)
Mean Stance Knee Rotation, Median (IQR)	-3 (-14–4)
Initial Contact Ankle Dorsiflexion, Median (IQR)	-2 (-7–3)
Mean Stance Ankle Dorsiflexion, Median (IQR)	6 (2–10)
Mean Swing Ankle Dorsiflexion, Median (IQR)	-3 (-10–3)
Mean Stance Foot Progression, Median (IQR)	-7 (-16–3)
Normalized Cadence, Median (IQR)	0.27 (0.24–0.29)
Normalized Step Length, Median (IQR)	1.32 (1.14–1.47)

<https://doi.org/10.1371/journal.pone.0270731.t004>

passive walking dynamics may also be the mechanism underlying the energy penalty caused by knee hyperextension.

After observing the minimum stance-phase knee flexion result, we hypothesized that this reflected the energetic penalty for extending a loaded and flexed knee and thus raising the body center-of-mass (recall that the initial contact position is accounted for in the model). Such a pattern is sometimes called “*jump gait*”. We explored this hypothesis by stratifying the minimum stance phase knee flexion plot by initial contact knee flexion, binned into mild, moderate, and severe categories (Fig 10). We see that, in support of our hypothesis, only individuals who land in severe crouch derive a net energy consumption benefit from remaining in crouch. Not all these individuals derive the net energy consumption benefit, indicating that additional causal factors play a role. The results of the model are suggestive of a mechanism but need to be tested experimentally or through simulations.

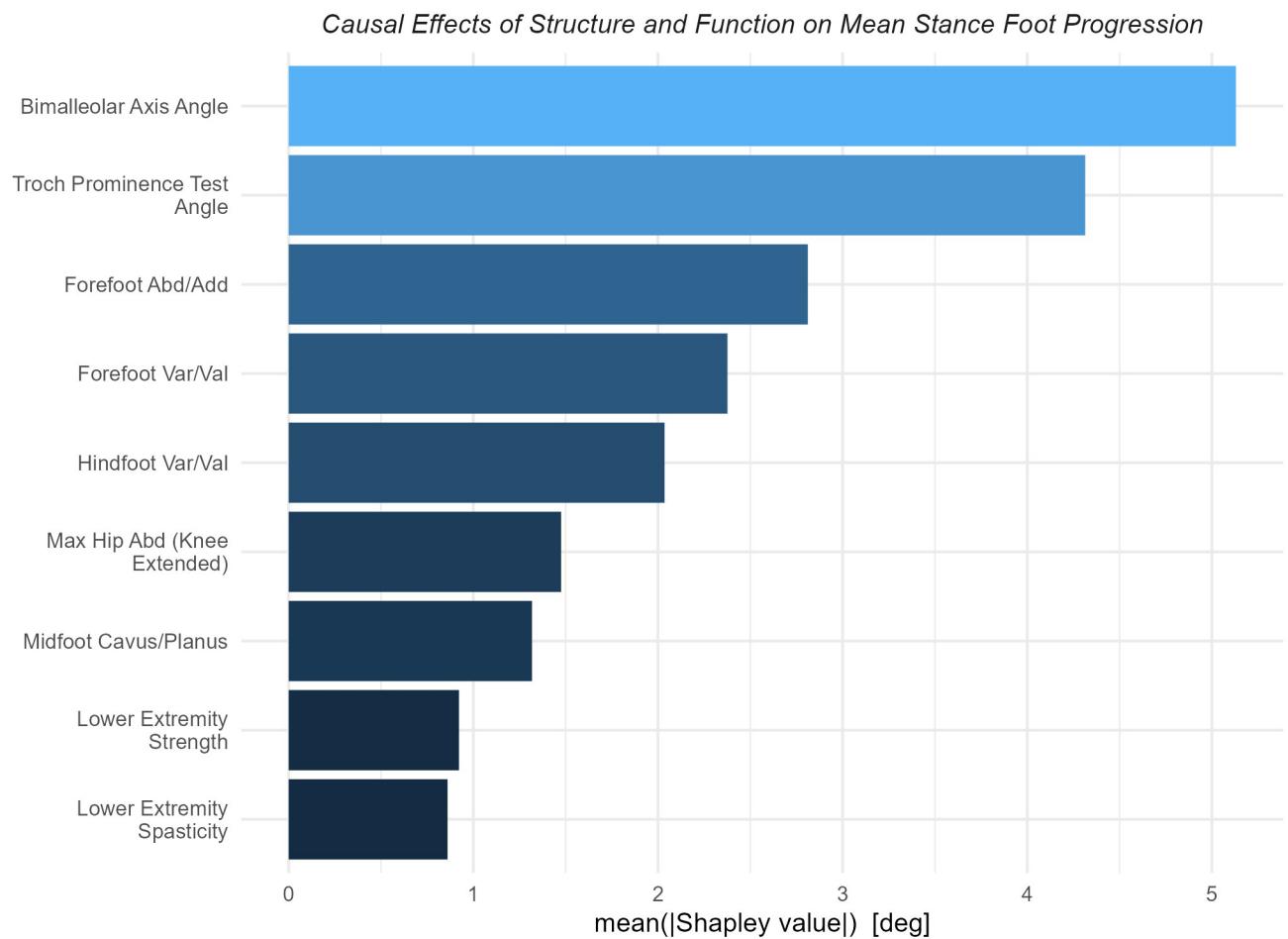
Two examples, each from individuals with high net energy consumption levels, show that different Gait Mechanics profiles can lead to similar levels of walking energy (Fig 11). The two cases have identical net energy consumption, but one is in severe crouch while the other is in equinus and knee hyperextension.

**Example 4: Direct effect of Age on net energy consumption.** Net energy consumption undergoes a substantial reduction with age (~ 1 SD) (Fig 12). For clarity we have fit a generalized growth (Richards') curve to the data [37]. This results in an initial net energy consumption of +0.49, final net energy consumption of -0.42, midpoint of maturation of 10.1 years, and time constant of 0.49. The age dependence shows significant energy decrease (0.91 SD) occurring primarily between ages 5–12 years.

**Table 5. Mobility, energy, quality of life.**

Variable	N	Value
Functional Assessment Questionnaire Transform (FAQt), Median (IQR)	5,498	55 (32–79)
Net Energy Consumption, Median (IQR)	7,672	2.81 (1.55–4.32)
Body Image and Self-Esteem, Median (IQR)	839	50 (38–58)
Activities of Daily Living and Independence, Median (IQR)	835	75 (57–91)

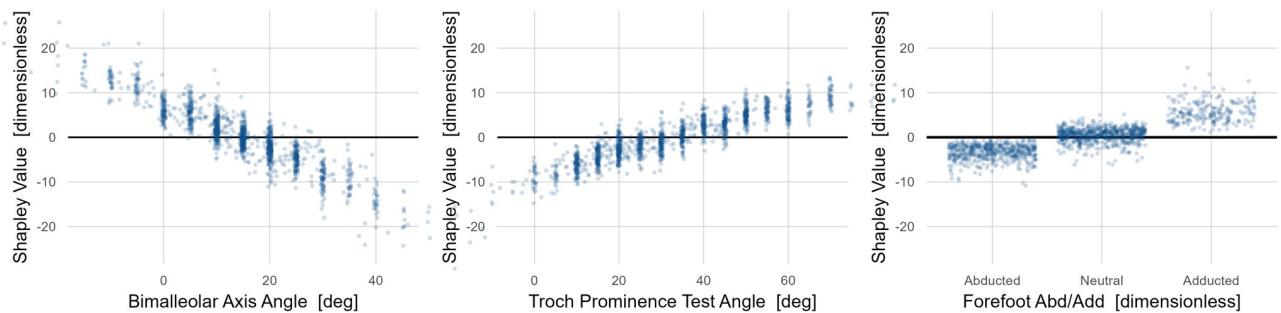
<https://doi.org/10.1371/journal.pone.0270731.t005>



**Fig 2. Importance of Structure and Function on mean stance foot progression.** The nine variables with the largest absolute Shapley values for mean stance foot progression are shown. Unsurprisingly, torsions of the tibia and femur are the two largest contributors to mean stance foot progression. Foot deformities comprise the next three largest causes.

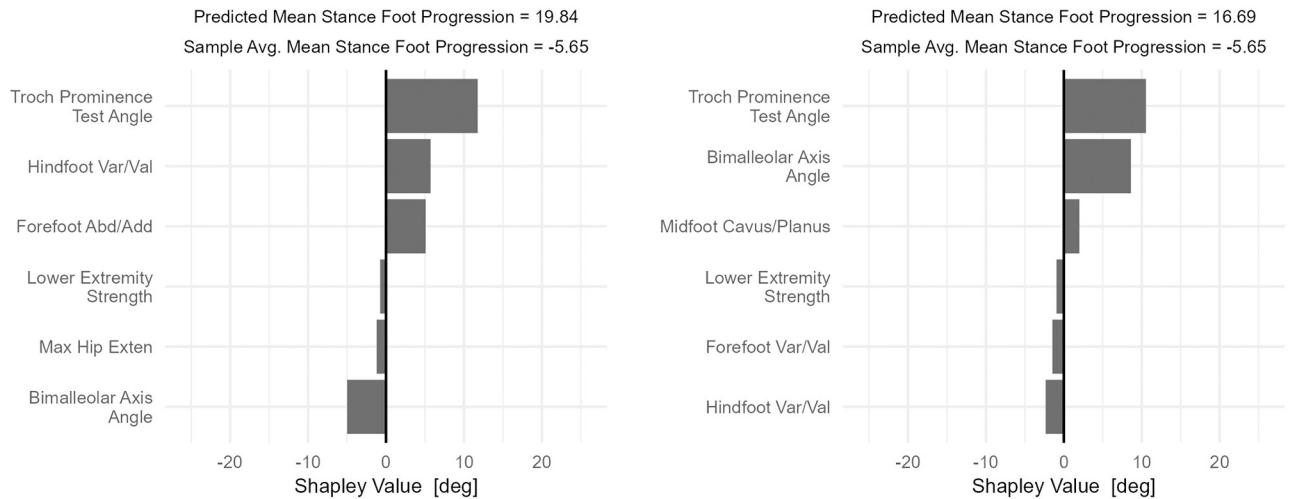
<https://doi.org/10.1371/journal.pone.0270731.g002>

An implication of this effect's importance can be seen in outcomes of selective dorsal rhizotomy. Elevated net energy consumption is commonly observed in children with cerebral palsy. Historically, selective dorsal rhizotomy was believed to be effective at reducing net energy consumption. Recent studies have shown that this is not true [38, 39]. At our center the



**Fig 3. Dependence plots for Structure and Function variables with large effects on mean stance foot progression.** The three most important causal variables are shown. The effects of long bone torsions (femur and tibia) are nearly linear. Note that for many years measurements were only recorded at five-degree increments, leading to vertical striations in the left and middle panels.

<https://doi.org/10.1371/journal.pone.0270731.g003>



**Fig 4. Case studies of Structure and Function effects on mean stance foot progression.** The top row shows patients with significant in-toeing. On the top-left, the primary cause is excess femoral anteversion, with additional contributions from hindfoot varus and forefoot adduction. On the top-right femoral anteversion and tibial torsion both contribute. The bottom row shows patients with significant out-toeing. On the bottom left, external tibial torsion and forefoot deformity cause the out-toeing. On the bottom right, both femoral retroversion and external tibial torsion contribute meaningfully, along with forefoot abduction.

<https://doi.org/10.1371/journal.pone.0270731.g004>

mean baseline and follow-up ages for selective dorsal rhizotomy evaluations are 6.3 and 7.8 years, which is during the period where age alone—Independent of treatment—causes rapid reductions in net energy consumption. Thus, the misattribution of age effects as treatment effects probably contributed to the erroneous ideas about selective dorsal rhizotomy and net energy consumption.

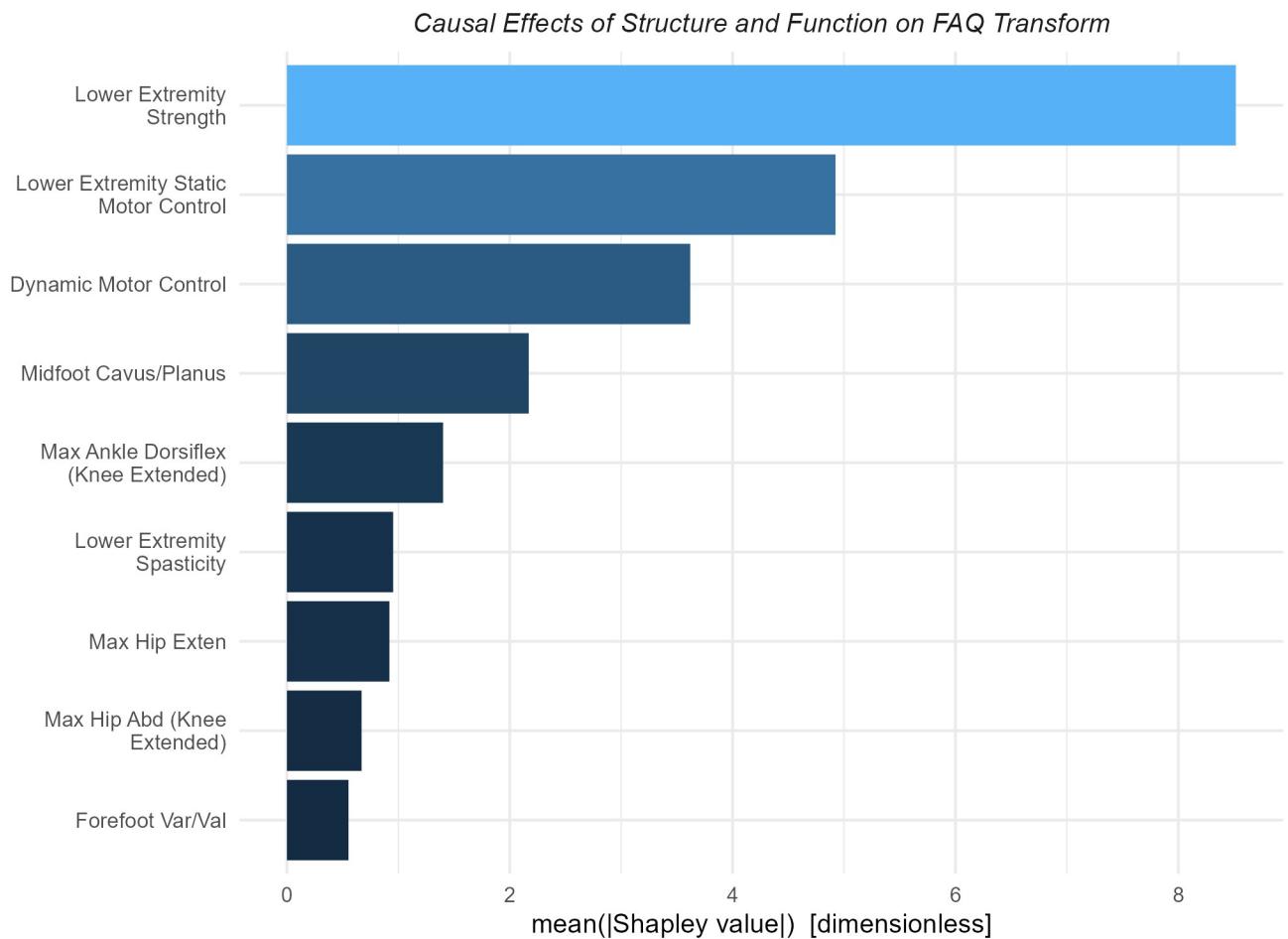
**Example 5: Total effects of (1) Structure and Function on Activities of Daily Living and Independence and (2) Gait Mechanics on Body Image and Self-Esteem.** Strength and motor control (static and dynamic) are the largest *Structure and Function* variables affecting Activities of Daily Living and Independence (Fig 13). As has been discussed earlier, these neurological factors are difficult to change with currently available treatments.

The next largest causal factors were related to ankle and foot deformity. Among *Gait Mechanics* variables affecting Body Image and Self-Esteem, large pelvic tilt range-of-motion, rapid cadence, and foot progression were the three largest causal factors. The model was able to identify detrimental effects on Body Image and Self-Esteem of both out- and in-toeing (external and internal foot progression)—with internal having a substantially larger impact (Fig 14). Foot progression is effectively treated with femoral and tibial derotation osteotomies and bony foot surgery.

## 4. Discussion

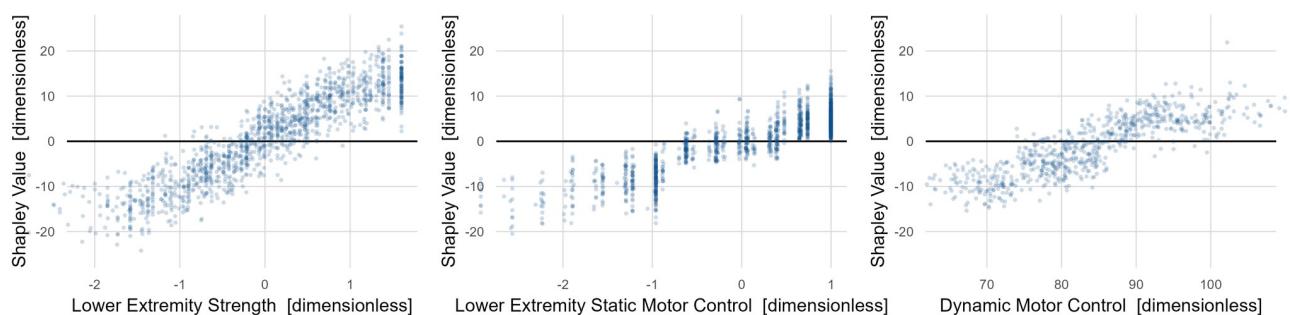
### 4.1 Summary

The purpose of this study was to propose a comprehensive model for evaluating causal effects of walking impairments and their consequences. We have described the model, demonstrated its plausibility using data from a pediatric clinical motion analysis laboratory, and presented several ways that it can be used to test and expand our understanding of walking impairments. The model provides results and data that can further our understanding of the relationships between treatments and outcomes by identifying the important sources of an impairment. The model's output reinforces some beliefs, provides useful details regarding others, and generates



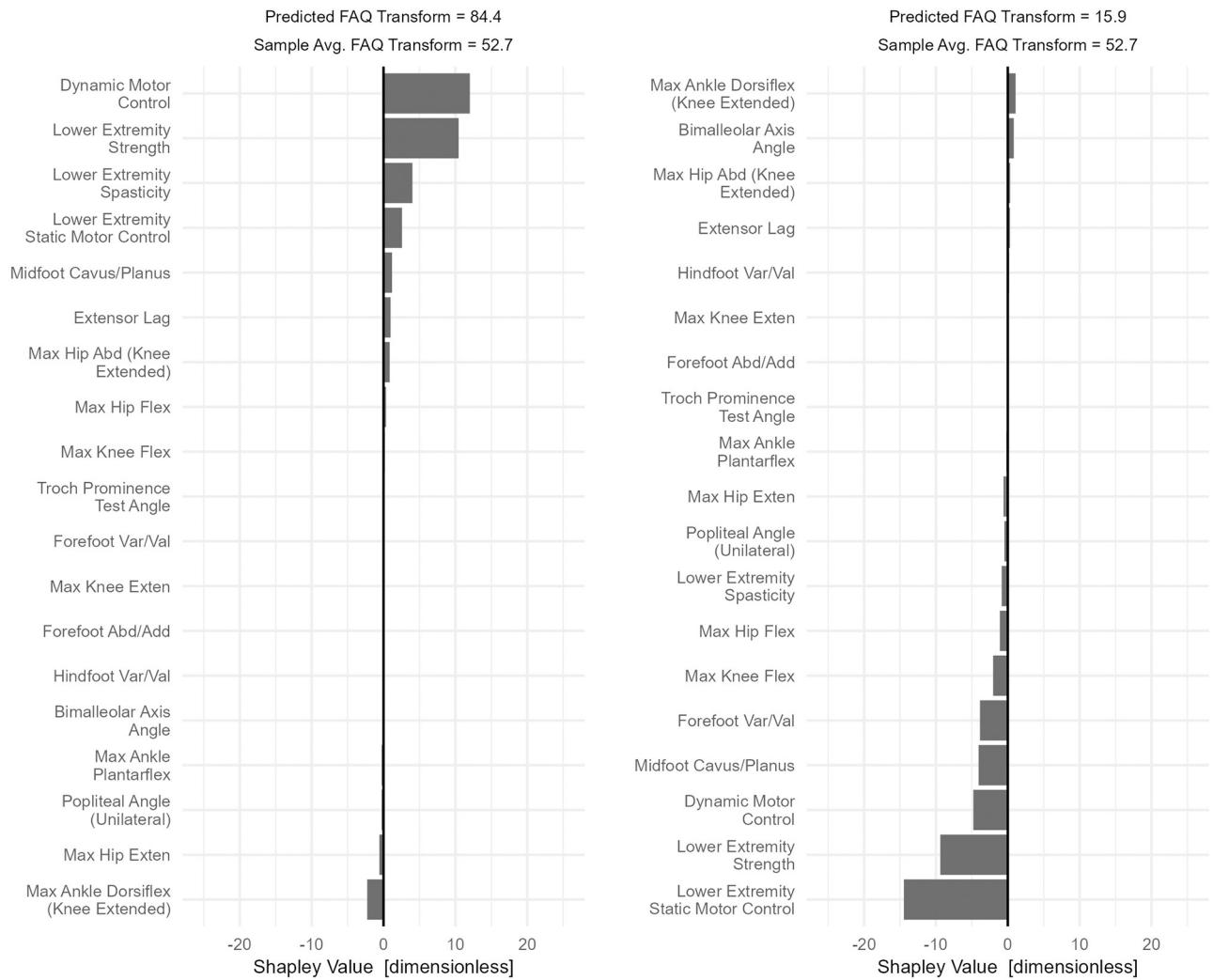
**Fig 5. Importance of Structure and Function on FAQt.** Strength and motor control (static and dynamic) are the main contributors to mobility as measured by the FAQt. These neurological variables have effects around two- to four-times larger than any other variable.

<https://doi.org/10.1371/journal.pone.0270731.g005>



**Fig 6. Dependence plots for Structure and Function variables with large effects on FAQt.** Strength and motor control (static and dynamic) constitute the top causal contributors to mobility as measured by the FAQt. The positive effect of dynamic motor control plateaus at around 90, which is 1 SD below typical.

<https://doi.org/10.1371/journal.pone.0270731.g006>



**Fig 7. Case studies of Structure and Function effects on the FAQt.** Good mobility (high FAQt) is caused by good strength and good motor control, poor mobility is caused by weakness and poor motor control. Strength and motor control are related, but that fact has been accounted for in the chosen adjustment set.

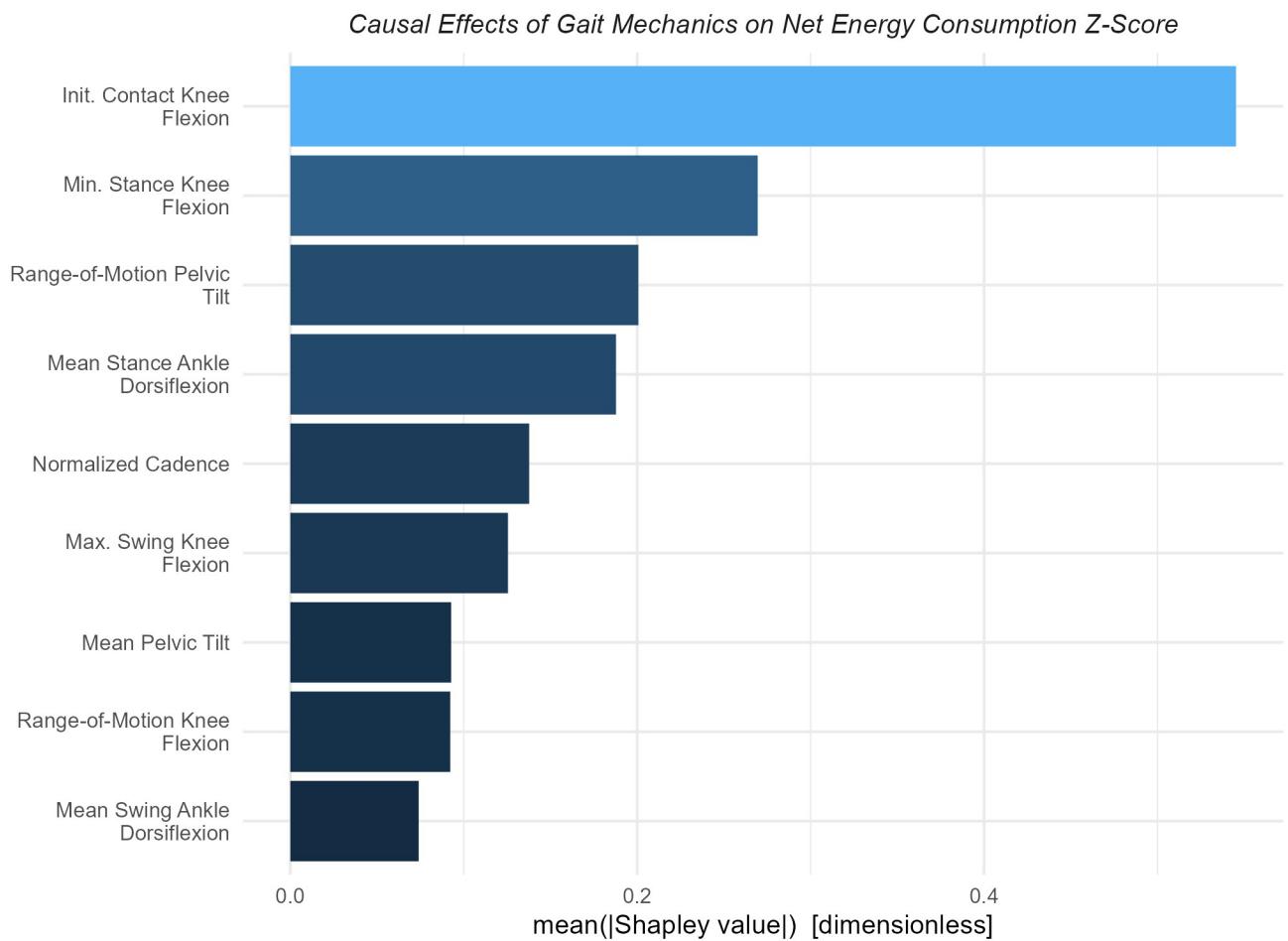
<https://doi.org/10.1371/journal.pone.0270731.g007>

new hypotheses. The model can also be used to estimate causal effects for individuals, with obvious implications for supporting treatment decisions.

## 4.2 Examples

This is not a “validation” study, *per se*, and the examples were not meant to prove the model is correct. Nevertheless, it is important that the model produces results that provide an element of face validity by reproducing—perhaps with more detail and rigor—existing knowledge. The large impact of femoral anteversion and tibial torsion on in- and out-toeing is not surprising. However, in addition to confirming this well-known effect, the model estimates the functional form, showing that tibial torsion is about twice as impactful as femoral anteversion on a degree-for-degree basis.

We can compare this result with clinical data. Consider two groups of limbs: One group underwent tibial derotation osteotomies without femoral derotation osteotomies or bony foot surgery (Tibial Derotation). The other group underwent femoral derotation osteotomies

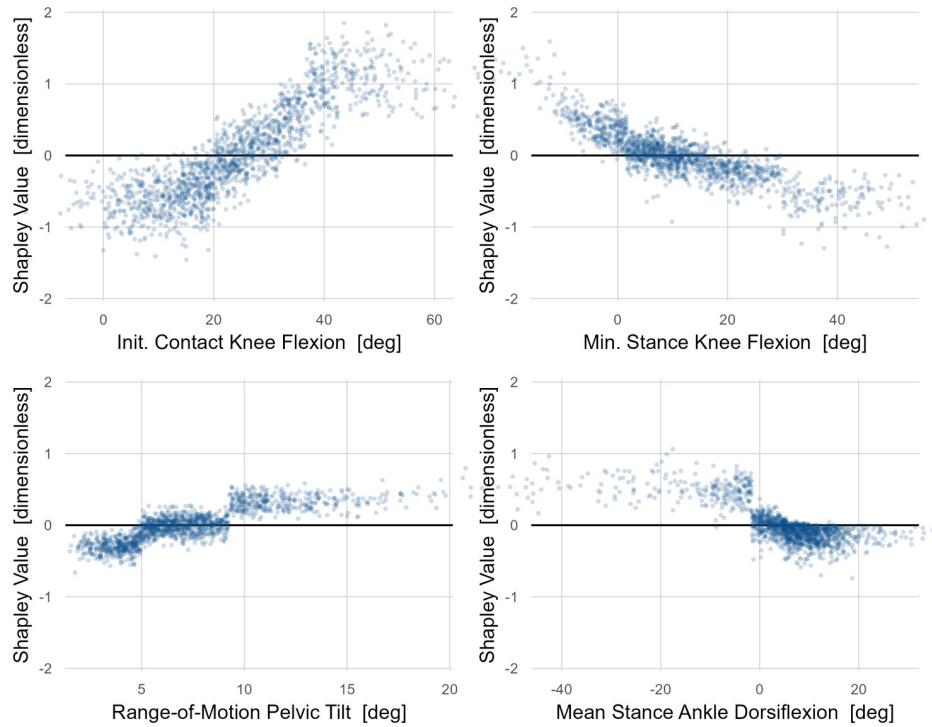


**Fig 8. Importance of Gait Mechanics on net energy consumption.** Knee flexion at initial contact has a causal impact twice as large as the next largest contributors, which are comprised of stance phase kinematic features of the knee and ankle.

<https://doi.org/10.1371/journal.pone.0270731.g008>

without tibial derotation osteotomies or bony foot surgery (Femoral Derotation). Examining the change in mean stance foot progression compared to the change in torsion (estimated from pre- and post-surgical physical examination) reveals slopes that closely match the causal prediction (Fig 15). Note that we do not expect exact matches due to additional factors that can influence foot progress, such as age, other treatments, and compensatory movements.

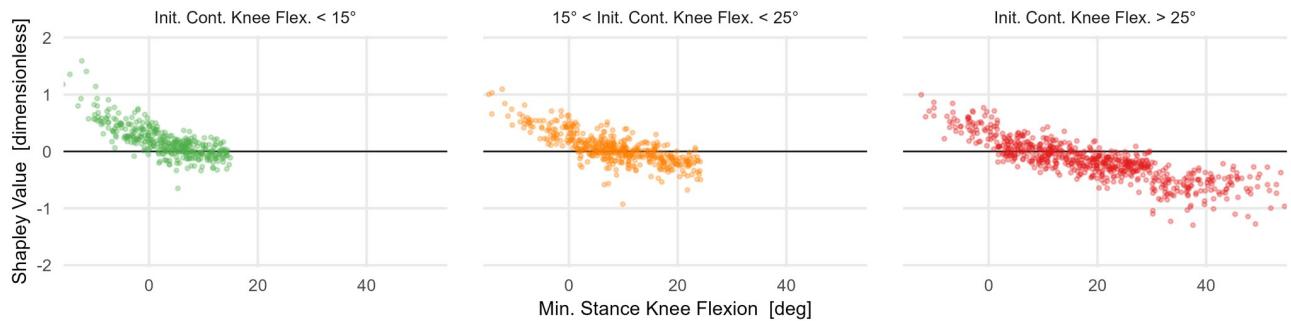
In- and out-toeing were also found to be detrimental to *Body Image and Self-Esteem*, a widely believed concept, but one that is not easily seen in raw data. In addition to clearly identifying this causal link, the model showed that in-toeing was around twice as impactful as out-toeing. Like the torsion findings, the direct effect of age on net energy consumption is not novel, but the model demonstrates this effect using substantially different methods than have been used in the past. Furthermore, we derive an explicit functional form for the age-dependence that can be used for adjusting observed treatment effects. The evaluation of net energy consumption also confirmed the critical role of knee flexion. However, rather than finding a monotonic crouch → energy consumption relationship, we saw a subtle yet sensible interaction effect between initial contact knee position, minimum stance-phase knee flexion, and



**Fig 9. Dependence plots for Gait Mechanics variables with large effects on net energy consumption.** Being flexed at initial contact imparts a substantial net energy consumption penalty. Knee hyperextension is costly, but maintaining a flexed position is not. This counterintuitive finding is discussed further in text. A large pelvis range-of-motion and stance phase ankle plantarflexion were also substantial causes of high net energy consumption—though with much smaller effects than knee kinematics.

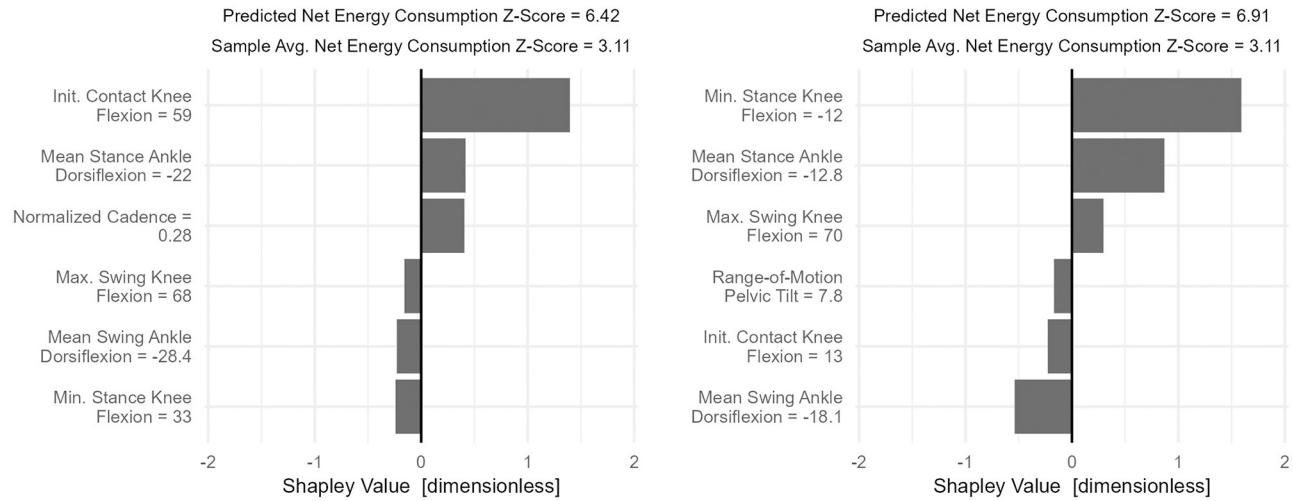
<https://doi.org/10.1371/journal.pone.0270731.g009>

energy consumption. While the purported mechanism needs to be tested, uncovering unknown or poorly understood phenomena like this exemplifies an important role that causal models can play. We found that strength and motor control are the key causal factors for both *Mobility* and *Quality-of-Life* (Activities of Daily Living and Independence). This result agrees with clinical beliefs and suggests why correction of orthopedic impairments does not lead, on average, to improvements in mobility [17].



**Fig 10. The effect of stance-phase knee flexion on net energy consumption, stratified by initial contact knee flexion.** The energy benefit of stance-phase crouch scales with the amount of initial contact knee flexion. Only individuals who land in a severely crouched position exhibit a large energy benefit from a large minimum stance-phase knee flexion. Not all individuals who land in severe crouch derive the benefit, suggesting a role for additional causal factors.

<https://doi.org/10.1371/journal.pone.0270731.g010>

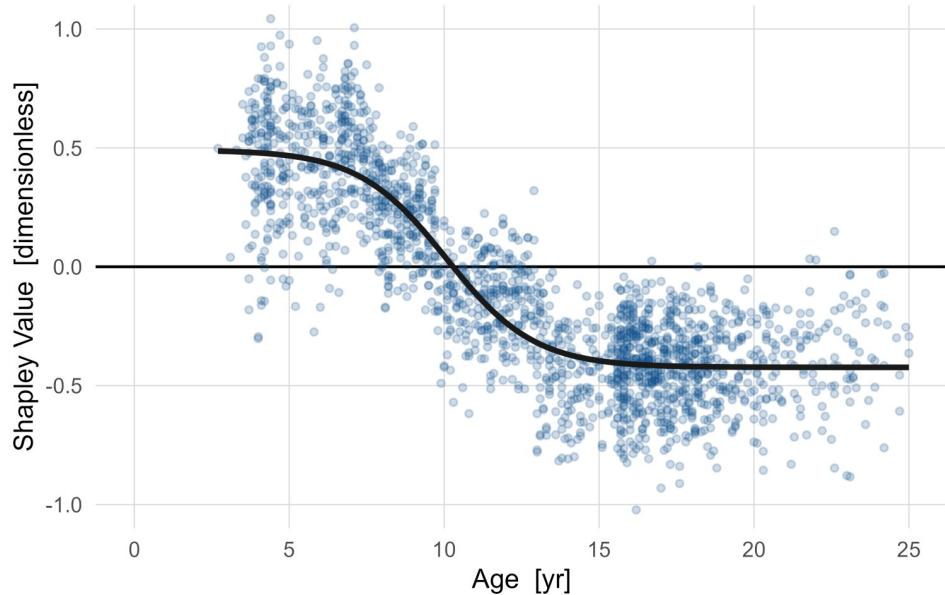


**Fig 11. Case studies of Gait Mechanics effects on net energy consumption.** The two individuals depicted have net energy consumption between 6 and 7 SD above speed-matched typically developing controls. For the individual on the left, initial contact knee flexion (59°) is an important cause. The individual is in severe crouch (minimum stance knee flexion = 33°)—yet this crouch lowers the net energy consumption. The hypothesized mechanism for this is described in the text. For the patient on the right, knee hyperextension and excessive ankle plantarflexion both contribute meaningfully to the elevated net energy consumption.

<https://doi.org/10.1371/journal.pone.0270731.g011>

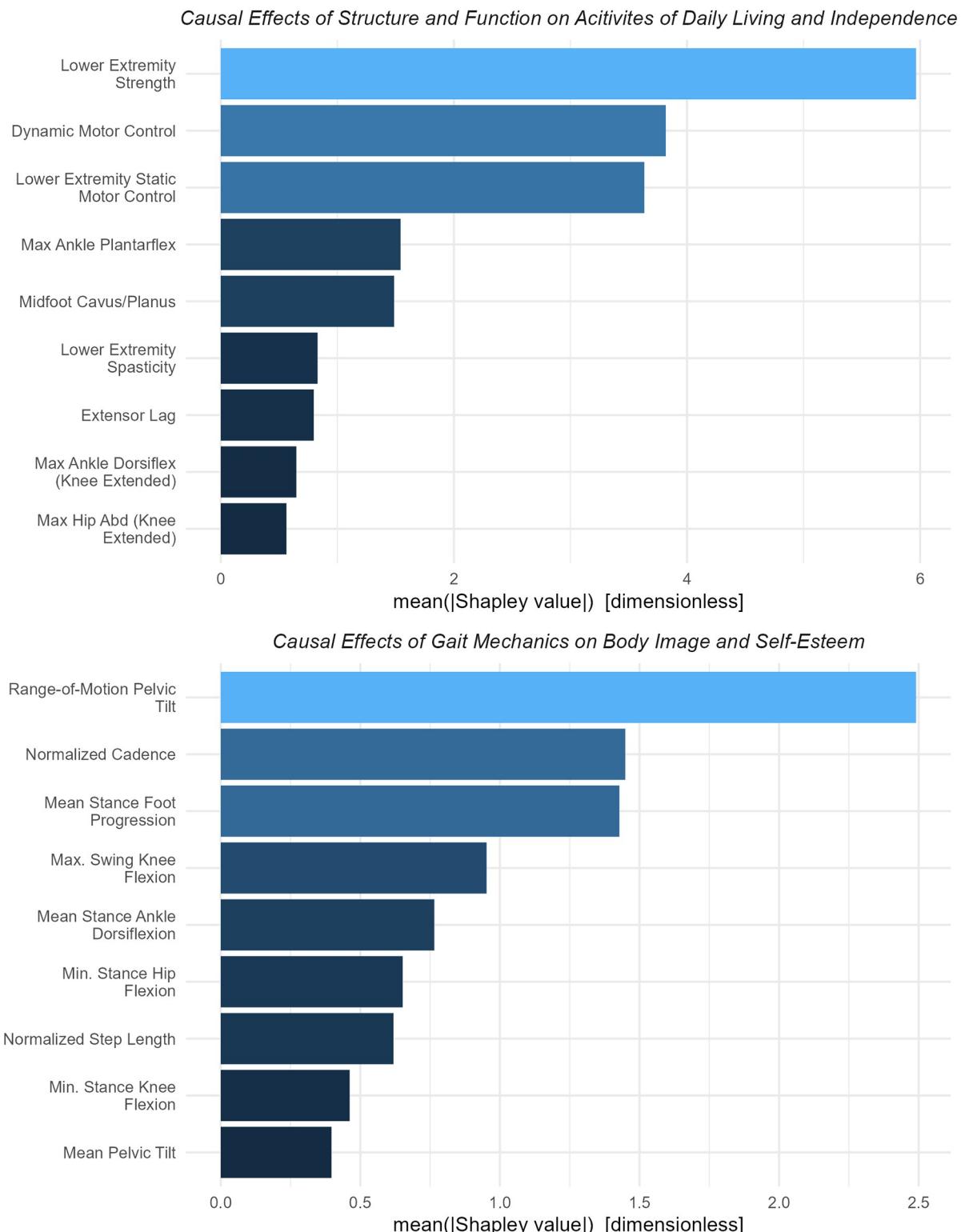
### 4.3 Interpreting model outputs

Shapley values are well-established, principled, and easily interpreted method for quantifying causal effects. Other commonly used methods are partial dependence plots, accumulated local effects plots, and individual conditional expectation plots. All these methods give more-or-less



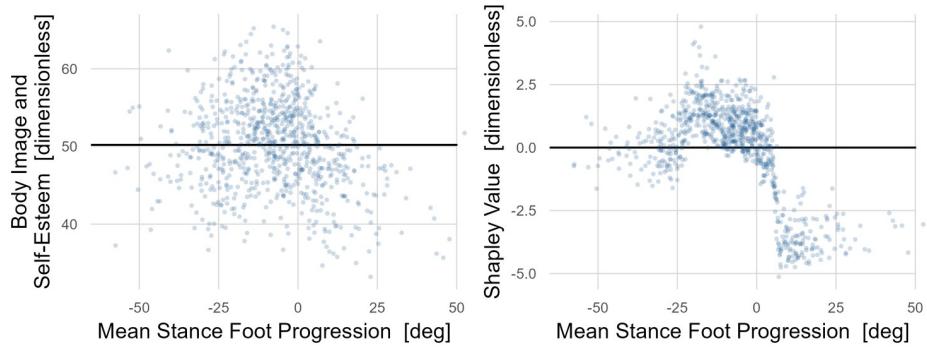
**Fig 12. The direct effect of age on net energy consumption.** Meaningful maturation effects accounting for around 1 SD of net energy consumption reduction can be clearly seen. The period of rapid reduction occurs from around 7.5–12.5 years. Evaluating the effect of treatment on energy consumption during this epoch must account for the direct effects of age to avoid conflating age-related reductions with treatment effects.

<https://doi.org/10.1371/journal.pone.0270731.g012>



**Fig 13. Importance of Structure and Function effects on Activities of Daily Living and Independence, Importance of Gait Mechanics effects on Body Image and Self-Esteem.** Strength and motor control dominate causes of Activities of Daily Living and Independence. These factors are difficult to change using currently available treatments. Some of the important causal variables can be addressed with well-established treatments. For example, ankle plantarflexion (Activities of Daily Living and Independence) or mean stance foot progression (Body Image and Self-Esteem).

<https://doi.org/10.1371/journal.pone.0270731.g013>

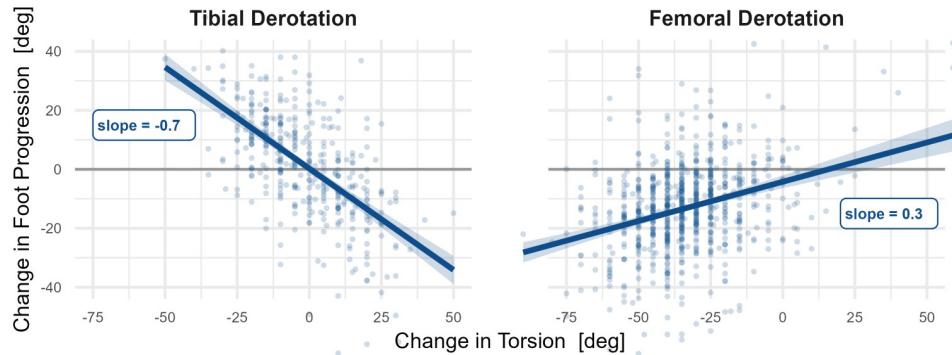


**Fig 14. The impact of in- and out-toeing on Body Image and Self Esteem.** Raw, non-causal (left) and causal (right) impacts of in- and out-toeing (mean stance foot progression) are shown. The raw data is widely dispersed but is suggestive of detrimental effects when deviating from typical foot alignment (around  $-10^\circ$ ). These effects are much clearer in the causal dependence plot. Deviations in either direction from typical result in lower *Body Image and Self-Esteem*, but the impact of in-toeing (positive) is about twice as large as that of out-toeing. There is a precipitous drop in *Body Image and Self Esteem* starting at around  $0^\circ$ , which is probably where in-toeing becomes easily discernable.

<https://doi.org/10.1371/journal.pone.0270731.g014>

the same answer. This raises an important point regarding the sort of precision we expect from causal models. Many are familiar with the saying “*measure with a micrometer and cut with a hacksaw*”. Here, we do not have that problem since we are essentially using a hacksaw for both measuring and cutting. In other words, our goal is not to identify effects down to the millimeter or degree, but to identify the main causal factors, rule out non-causes unjustified beliefs, and derive reasonable sign and magnitude estimates for causal effect sizes [40].

Shapley values estimate causal factors and therefore give useful guidance for treatment decisions. However, Shapley values do not necessarily predict changes associated with treatment. There are several reasons for this. Multiple variables may change with a treatment, leading to a different causal impact due to interactions. Furthermore, patients can compensate for changes introduced by treatment. For example, suppose a patient’s Shapley profile indicated in-toeing largely caused by femoral anteversion. Now suppose the patient underwent a femoral derotation osteotomy of magnitude matching the Shapley value. Finally, suppose the patient compensated after surgery by internally rotating their pelvis by half the amount of the derotation. In this case, the treatment outcome would only be half of the original causal effect of anteversion, despite a full orthopedic correction. Additional work is needed to quantify the relationships between causal effects and treatment outcomes.



**Fig 15. Clinical estimates for the causal effects of tibial torsion and femoral anteversion on in- and out-toeing.** The effects are approximately linear, with slopes of  $-0.7$  and  $0.3$  for tibial and femoral torsion, respectively. This is similar to the predicted causal effect slopes of  $-0.6$  and  $0.3$  degrees.

<https://doi.org/10.1371/journal.pone.0270731.g015>

#### 4.4 Limitations

All causal models are proposals that can never be fully confirmed or refuted, and thus need to be evaluated critically. The minimal level of model checking involves testing implied conditional independencies to establish model plausibility. Most implied conditional independencies could not be evaluated for the proposed model due to the large numbers of conditioning variables. Those that could be tested supported the plausibility of the model. We know that there are missing paths in the model. Some of these missing paths could represent important causes while others could introduce bias. However, not every missing path leads to bias. For this to be the case, missing variables would, at a minimum, need to be a common cause of the exposure and outcome of interest. For example, cognition would need to affect *Structure and Function* and *Gait Mechanics* for its absence to bias the findings regarding torsion and in-toeing. Furthermore, even if biasing paths exist, it is not clear how large an effect such a bias would have. This question can be studied with sensitivity analysis.

The *Diagnosis* node is crude. The most obvious next element to add to this node is a measure of severity. Those involved in CP research will immediately notice the absence of the Gross Motor Function Classification System (GMFCS), which is often viewed—explicitly and implicitly—as a severity measure. We have left the GMFCS out of the model for two main reasons. The first reason is that we do not have similar scales for other diagnoses and would thus introduce significant imbalance in missing values. This is a relatively minor problem since missing values can be handled seamlessly by *bartMachine*. The second reason, which is more important and more intriguing, is that it is unclear whether the GMFCS level is a cause or an effect. The original description suggests GMFCS as a cause—a fixed value that is a surrogate for severity [41, 42]. However, in practice, a GMFCS level is assigned based on observed functional skills like climbing stairs or carrying objects. This creates a paradox, since the skills dictating the GMFCS assignment are caused by *Structure and Function* and *Gait Mechanics* impairments. In other words, it is not clear whether the causal arrow is GMFCS → *Structure and Function/Gait Mechanics* or *Structure and Function/Gait Mechanics* → GMFCS. Given the lack of formal causal modeling in CP research, it is not surprising that this paradox has not been discussed previously. The proper causal model would appear to at least contain GMFCS ← Severity → *Structure and Function/Gait Mechanics*, with Severity a latent variable. We have experimented with other severity measures such as age of walking onset, which seems to work well. However, establishing the validity of such a severity measure is beyond the scope of this study.

Our study is limited by sample bias. This is an observational study, and our data comes from a convenience sample of individuals seen for clinical necessity. One obvious problem with such a sample is that it reflects individuals with larger impairments than the population. More subtle biases may arise from non-representative cultural and socioeconomic factors. Another consequence of our convenience sample is that it is clearly biased towards younger individuals (Age, Median (IQR) = 10.5 (7.7–13.8)). The factors that impair gait in older individuals may vary substantially from those represented in these data. It is critical to recognize the limited applicability of this model to an ageing population.

In this study we have only examined continuous outcomes. The basic methodology applies directly to categorical outcomes (binary, ordered, polytomous), but for ordered and polytomous outcomes it is a necessary and straightforward modification to use a different BART engine, since *bartMachine* does not handle ordered and polytomous outcomes. We have successfully used the *BART* package for these types of outcomes [30].

#### 4.5 Future directions

As stated at the outset, this model is a proposal that is meant to be critically evaluated, validated or refuted, altered, and improved over time. Such improvements might include the introduction of new nodes, variables, and paths. We are particularly interested in modeling the effects of social determinants of health, though we recognize the monumental challenge this presents. *Quality-of-Life* is affected by many factors not in the model, such as employment, social relations, and other health conditions, to name a few. Nevertheless, the results (sensibly) suggest that certain impairments of *Structure and Function* and *Gait Mechanics* reduce *Quality-of-Life*. The current model does not include pain. Pain is clearly an important phenomenon, but it is unclear how to properly include it. For example, pain can be a cause of *Gait Mechanics* (antalgic gait caused by osteoarthritic pain) or pain can be an effect of *Gait Mechanics* (e.g., knee pain caused by pre-patellar pressures present in crouch gait).

The replication crisis in science is real. There are vanishingly few examples of important clinical gait studies being replicated or refuted [43]. To encourage open and transparent scientific practice, we are providing a detailed *Rmarkdown* file and a sample dataset. We expect that this will greatly ease the burden of implementing this model in other centers and will help researchers identify model shortcomings and propose possible improvements.

Causal inference is inherently ambiguous since we cannot observe multiple realizations of the same person with different characteristics. Causal models must be evaluated through indirect means and reasoning [44]. We have proposed a model, demonstrated its utility, and provided limited face validity. We have made our methods transparent and easy to implement and scrutinize by peers. It is important to design experiments and simulations that test the predictions and underlying assumptions of this model. That difficult task must be the source of future efforts.

### Supporting information

**S1 File. Causal model–Rmarkdown.** This compressed folder contains the files necessary to generate the model and output from this manuscript. The file *ReadMe.txt* contains basic instructions (knowledge of *R* is assumed). The file *Causal Model.Rproj* is the *R* project file that should be launched. The file *Main.Rmd* is the *Rmarkdown* file that builds the model and generates the output. The file *sample-data.RData* contains deidentified data needed as input to the model.

(ZIP)

**S1 Appendix.**

(DOCX)

**S2 Appendix.**

(DOCX)

**S3 Appendix.**

(DOCX)

### Author Contributions

**Conceptualization:** Michael H. Schwartz.

**Data curation:** Michael H. Schwartz, Andrew J. Ries.

**Formal analysis:** Michael H. Schwartz.

**Investigation:** Michael H. Schwartz.

**Methodology:** Michael H. Schwartz, Katherine M. Steele.

**Project administration:** Michael H. Schwartz.

**Resources:** Michael H. Schwartz, Andrew G. Georgiadis.

**Software:** Michael H. Schwartz, Bruce A. MacWilliams.

**Supervision:** Michael H. Schwartz.

**Validation:** Michael H. Schwartz.

**Visualization:** Michael H. Schwartz.

**Writing – original draft:** Michael H. Schwartz.

**Writing – review & editing:** Michael H. Schwartz, Katherine M. Steele, Andrew J. Ries, Andrew G. Georgiadis, Bruce A. MacWilliams.

## References

1. Ethgen O, Bruyère O, Richy F, Dardennes C, Reginster J-Y. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am.* 2004; 86: 963–974. <https://doi.org/10.2106/00004623-200405000-00012> PMID: 15118039
2. Rozumalski A, Galarraga O, Schwartz M, Desailly E. Experience in a multinational collaboration to predict kinematic outcomes after SEMLS in children with cerebral palsy. 2019. <https://doi.org/10.1016/j.gaitpost.2019.07.119>
3. Hill J, Stuart EA. Causal Inference: Overview. In: Wright JD, editor. International Encyclopedia of the Social & Behavioral Sciences ( Second Edition). Oxford: Elsevier; 2015. pp. 255–260.
4. app.dimensions.ai. 10 May 2022 [cited 10 May 2022]. [https://app.dimensions.ai/discover/publication?search\\_mode=content&search\\_text=causal%20inference&search\\_type=kws&search\\_field=full\\_search](https://app.dimensions.ai/discover/publication?search_mode=content&search_text=causal%20inference&search_type=kws&search_field=full_search)
5. Pearl J. Causal diagrams for empirical research. *Biometrika.* 1995; 82: 669–688. <https://doi.org/10.1093/biomet/82.4.669>
6. Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology.* 2013; 177: 292–298. <https://doi.org/10.1093/aje/kws412> PMID: 23371353
7. Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. *Ann Appl Stat.* 2010; 4: 266–298. <https://doi.org/10.1214/09-AOAS285>
8. Hill J. Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics.* 2011; 20: 217–240. <https://doi.org/10.1198/jcgs.2010.08162>
9. Dorie V, Hill J, Shalit U, Scott M, Cervone D. Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition. *Statist Sci.* 2019; 34: 43–68. <https://doi.org/10.1214/18-STS667>
10. Shapley LS. Notes on the n-Person Game—II: The Value of an n-Person Game. RAND Corporation; 1951. [https://www.rand.org/content/dam/rand/pubs/research\\_memoranda/2008/RM670.pdf](https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM670.pdf)
11. International Classification of Functioning, Disability and Health (ICF). 25 May 2022 [cited 25 May 2022]. <https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health>
12. Schwartz MH, Ries AJ, Georgiadis AG. Estimating the Efficacy of Common Treatments in Children and Young Adults Diagnosed with Cerebral Palsy Using Three Machine Learning Algorithms. *medRxiv;* 2021. p. 2021.10.06.21264624. <https://doi.org/10.1101/2021.10.06.21264624>
13. Kuo AD, Donelan JM. Dynamic Principles of Gait and Their Clinical Implications. *Phys Ther.* 2010; 90: 157–174. <https://doi.org/10.2522/ptj.20090125> PMID: 20023002
14. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. *Journal of Personality Assessment.* 1985; 49: 71–75. [https://doi.org/10.1207/s15327752jpa4901\\_13](https://doi.org/10.1207/s15327752jpa4901_13) PMID: 16367493
15. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Soc Sci Med.* 1998; 46: 1569–1585. [https://doi.org/10.1016/s0277-9536\(98\)00009-4](https://doi.org/10.1016/s0277-9536(98)00009-4) PMID: 9672396

16. Rosenbaum PR. Sensitivity to Hidden Bias. In: Rosenbaum PR, editor. *Observational Studies*. New York, NY: Springer; 2002. pp. 105–170.
17. Schwartz MH, Ries AJ, Georgiadis AG. Short-term causal effects of common treatments in ambulatory children and young adults with cerebral palsy: three machine learning estimates. *Sci Rep*. 2022; 12: 1–13. <https://doi.org/10.1038/s41598-022-11875-5> PMID: 35551496
18. Ashworth B. Preliminary trial of carisoprodol in multiple sclerosis. *Practitioner*. 1964; 192: 540–542. PMID: 14143329
19. Kendall HO, Kendall FP, Wadsworth GE. Muscles, Testing and Function. *American Journal of Physical Medicine & Rehabilitation*. 1973; 52: 43.
20. Steele KM, Rozumalski A, Schwartz MH. Muscle synergies and complexity of neuromuscular control during gait in cerebral palsy. *Developmental Medicine & Child Neurology*. 2015; 57: 1176–1182. <https://doi.org/10.1111/dmcn.12826> PMID: 26084733
21. Ehrig RM, Taylor WR, Duda GN, Heller MO. A survey of formal methods for determining functional joint axes. *J Biomech*. 2007; 40: 2150–2157. <https://doi.org/10.1016/j.jbiomech.2006.10.026> PMID: 17169365
22. Ehrig RM, Taylor WR, Duda GN, Heller MO. A survey of formal methods for determining the centre of rotation of ball joints. *J Biomech*. 2006; 39: 2798–2809. <https://doi.org/10.1016/j.jbiomech.2005.10.002> PMID: 16293257
23. Hof AL. Scaling gait data to body size. *Gait & Posture*. 1996; 4: 222–223. [https://doi.org/10.1016/0966-6362\(95\)01057-2](https://doi.org/10.1016/0966-6362(95)01057-2)
24. Schwartz MH, Koop SE, Bourke JL, Baker R. A nondimensional normalization scheme for oxygen utilization data. *Gait Posture*. 2006; 24: 14–22. <https://doi.org/10.1016/j.gaitpost.2005.06.014> PMID: 16112865
25. Schwartz MH, Aldahondo N, MacWilliams BA. A Patient-Reported Measure of Locomotor Function Derived from the Functional Assessment Questionnaire. *medRxiv*. 2021; 2021.06.12.21258826. <https://doi.org/10.1101/2021.06.12.21258826>
26. Thomason P, Tan A, Donnan A, Rodda J, Graham HK, Narayanan U. The Gait Outcomes Assessment List (GOAL): validation of a new assessment of gait function for children with cerebral palsy. *Developmental Medicine & Child Neurology*. 2018; 60: 618–623. <https://doi.org/10.1111/dmcn.13722> PMID: 29573409
27. Textor J, Zander B van der, Ankan A. *dagitty: Graphical Analysis of Structural Causal Models*. 2020. <https://CRAN.R-project.org/package=dagitty>
28. Kapelner A, Bleich J. *bartMachine: Machine Learning with Bayesian Additive Regression Trees*. *Journal of Statistical Software*. 2016; 70: 1–40. <https://doi.org/10.18637/jss.v070.i04>
29. Kapelner A, Bleich J. Prediction with missing data via Bayesian Additive Regression Trees. *Canadian Journal of Statistics*. 2015; 43: 224–239. <https://doi.org/10.1002/cjs.11248>
30. Sparapani R, Spanbauer C, McCulloch R. Nonparametric Machine Learning and Efficient Computation with Bayesian Additive Regression Trees: The BART R Package. *Journal of Statistical Software*. 2021; 97: 1–66. <https://doi.org/10.18637/jss.v097.i01>
31. Greenwell B. *fastshap: Fast Approximate Shapley Values*. 2021. <https://CRAN.R-project.org/package=fastshap>
32. Ankan A, Wortel IMN, Textor J. Testing Graphical Causal Models Using the R Package “dagitty”. *Current Protocols*. 2021; 1: e45. <https://doi.org/10.1002/cpz1.45> PMID: 33592130
33. Rosseel Y. *lavaan: An R Package for Structural Equation Modeling*. *Journal of Statistical Software*. 2012; 48: 1–36. <https://doi.org/10.18637/jss.v048.i02>
34. Gill PK, Steele KM, Donelan JM, Schwartz MH. Causal Effects Contributing to Elevated Metabolic Power During Walking in Children Diagnosed with Cerebral Palsy. *medRxiv*; 2022. p. 2022.01.26.22269878. <https://doi.org/10.1101/2022.01.26.22269878>
35. Duffy CM, Hill AE, Graham HK. The influence of flexed-knee gait on the energy cost of walking in children. *Dev Med Child Neurol*. 1997; 39: 234–238. <https://doi.org/10.1111/j.1469-8749.1997.tb07417.x> PMID: 9183261
36. Steele KM, Shuman BR, Schwartz MH. Crouch severity is a poor predictor of elevated oxygen consumption in cerebral palsy. *J Biomech*. 2017; 60: 170–174. <https://doi.org/10.1016/j.jbiomech.2017.06.036> PMID: 28734543
37. Richards FJ. A Flexible Growth Function for Empirical Use. *J Exp Bot*. 1959; 10: 290–301. <https://doi.org/10.1093/jxb/10.2.290>
38. MacWilliams BA, McMulin ML, Duffy EA, Munger ME, Chen BP-J, Novacheck TF, et al. Long-term effects of spasticity treatment, including selective dorsal rhizotomy, for individuals with cerebral palsy.

Developmental Medicine & Child Neurology. 2021;n/a. <https://doi.org/10.1111/dmcn.15075> PMID: 34755903

- 39. Zaino NL, Steele KM, Donelan JM, Schwartz MH. Energy consumption does not change after selective dorsal rhizotomy in children with spastic cerebral palsy. *Dev Med Child Neurol.* 2020; 62: 1047–1053. <https://doi.org/10.1111/dmcn.14541> PMID: 32306392
- 40. Gelman A, Carlin J. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspect Psychol Sci.* 2014; 9: 641–651. <https://doi.org/10.1177/1745691614551642> PMID: 26186114
- 41. Rosenbaum PL, Palisano RJ, Bartlett DJ, Galuppi BE, Russell DJ. Development of the Gross Motor Function Classification System for cerebral palsy. *Dev Med Child Neurol.* 2008; 50: 249–253. <https://doi.org/10.1111/j.1469-8749.2008.02045.x> PMID: 18318732
- 42. Rosenbaum PL, Walter SD, Hanna SE, Palisano RJ, Russell DJ, Raina P, et al. Prognosis for gross motor function in cerebral palsy: creation of motor development curves. *JAMA.* 2002; 288: 1357–1363. <https://doi.org/10.1001/jama.288.11.1357> PMID: 12234229
- 43. Schwartz MH, Ries AJ. Rectus femoris transfer in children with cerebral palsy: comparing a propensity score-matched observational study to a randomized controlled trial. *Developmental Medicine & Child Neurology.* 2021; 63: 196–203. <https://doi.org/10.1111/dmcn.14709> PMID: 33084049
- 44. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med.* 1965; 58: 295–300. PMID: 14283879