Representing Progression and Interactions of Comorbidities in Aggregate and Individual-Based Systems Models

Nathaniel Osgood
Department of Computer Science
University of Saskatchewan
osgood@cs.usask.ca

Abstract: Health policy models have attracted significant offered important insights in to health trends and policy selection. More complete accounting for the cost and health implications of upstream interventions is hindered by the need to consider impact on, and interactions between, multiple comorbidities. Within this paper, we explore several distinct approaches for representing comorbidities, some of them at the aggregate level, and some of them at the individual level. All of these representations have the virtue of being declarative, in that they allow the user to focus on *what* is to be characterized, rather than how it is to be implemented. Our exploration suggests that while several aggregate representations of comorbidities are possible, they suffer from a variety of shortcomings, ranging from low fidelity to combinatorial blowup. While individual-level representations impose a heavy performance load, greater difficulties in calibration and less rapid analysis, such representations do offer greater transparency, modifiability, scalability, and modularity, and ease of representing transmission and influence networks. With much to recommend each approach, further research is needed to shed additional light on the tradeoffs and identify situations where one representation is preferable to another.

1 Introduction

It is well known that illnesses cluster in individuals, and in disadvantaged population groups [2]. For example, the constellation of T2DM and its comorbidities cluster tightly enough to have been granted their own term: "Metabolic Syndrome" (also called "dysmetabolic syndrome" or "syndrome X") [11]. The reasons for such clustering include shared risk factors (for example, the risks of both coronary artery disease and T2DM are both elevated by obesity and dietary imbalances), etiological linkages in disease progression [12], environmental factors (e.g. characteristics of the built environment), access to healthy foods [13], health care, education, etc. In the communicable disease area, comorbid conditions are common. For some chronic diseases, large numbers of comorbidities are possible. For example, in diabetes, high blood glucose levels during T2DM can worsen coronary artery disease, contribute to kidney dysfunction, nerve damage, retinal damage, and other conditions. But co-morbidities play an important role in the infectious disease area as well. For example, individuals with one type of sexually transmitted infection are at high risk of co-infection with other strains or other STIs, due both to common risk factors and to some biological mechanisms that enhance risk of infection. Blood-borne pathogens such as Human Immunodeficiency Virus and Hepatitis C are transmitted by similar risky practices (e.g. needle sharing during intravenous drug use), and then interact both biologically and through restrictions of treatment regimen. Finally, because of the adverse immune impacts of many chronic diseases (e.g.

diabetes, chronic kidney disease) and risk factors (smoking, obesity) those with one set of chronic conditions often suffer from infection as well.

While there have been considerable success stories in modeling the health and cost implications of health interventions for specific conditions (for example, diabetes [3-5] [6, 7] [8] [9, 10]), the tight coupling between different conditions makes it desirable to examine simultaneous policy impact on broad sets of conditions. At the same time, the representation of multiple co-morbidities – like the representation of other heterogeneities – can add considerable complexity to the modeling process. This paper seeks to explore some options for representing such co-morbidities, as well as relevant risk factors.

This paper is organized as follows. The next section of the paper provides information on the importance of comorbidities, introducing diabetes as an example. Section 3 examines two techniques for representing comorbidities in an aggregate model, based (in one case) on statistical correlations between comorbidities and stratification of the population by stages of progression (on the other). We also discuss co-flows and attribute-based disaggregation. Motivated by some of the difficulties of these representations, Section 4 discusses individual-based representations – representations in which the population is modeled as a collection of simulated individuals, within each of which we model the dynamics of disease progression. These representations are notable in that they are declarative – describing the *what* of disease progression rather than the how. As such, they represent departures from the historical tendency for agent behavior to be described using general-purpose programming languages. In light of the expressiveness of individual-based models for representing comorbidities, Section 5 examines some other advantages of such models, namely the ability to more expressively model inter-person health linkages. Finally, we conclude the paper with a brief summary.

2 Comorbidities and Their Importance

For several reasons, the existence of comorbidities can significantly impact the relative cost and health impact of interventions.

At the most basic level, distinct comorbidities act as competing risks. As a result, downstream interventions that address one comorbidity without reducing the severity of the others will offer more limited benefit in terms of quality of life and mortality than one might otherwise expect [14]. For example, while the elevated blood sugars seen during diabetes do damage heart tissue, some components of the association between diabetes and heart disease are non-causal. If attributable mortality statistics indicate that 18% of individuals die prematurely of heart disease, even a downstream “silver bullet” that eliminates heart disease may not reduce premature mortality by 18%, as the individuals who would have died of heart disease may well die of other common comorbidities that tend to exist in the same individuals [15].

The presence of comorbidities can also lead to unexpected windfalls in the impact of upstream interventions. For example, the fact that T2DM and coronary artery disease cluster within individuals can boost the efficacy of regular physician checkups on obese patients. Similarly, because chronic comorbidities share upstream population health determinants, effective upstream interventions can often help lower the likelihood and severity of many downstream conditions. For example, physician screening

or counseling of obese patients on the importance of nutrition or exercise can reduce the risk or severity of T2DM, coronary artery disease, and congestive heart failure.

Windfalls can also result when causal linkages exist between the progression of comorbidities, as can also lead to strengthened interventions, reflecting the fact that slowing the progression of one comorbidity may also hinder the progression of others. For example, effective control of blood glucose levels can not only slow progression of diabetes, but can also lessen the progression of coronary artery disease. Such effects suggest that the impact of a given intervention may be greater than a naïve (static) competing risks approach might suggest.

Finally, the presence of comorbidities may have a significant impact on Health Related Quality of Life (HRQoL). Taking such comorbidities effects on HRQoL into account could be important when studying intervention impact.

Taken together, these considerations have significant implications for the health and cost impacts of interventions. Because of their implications for multiple downstream conditions, upstream interventions (e.g. investment in community recreation facilities and fostering more walkable neighborhoods) can offer significantly greater leverage (and correspondingly higher cost-effectiveness) than a disease-specific analysis would indicate: A single physician's time may eventually go a long way in preventing or reducing the severity of several conditions. By contrast, downstream interventions (e.g. a kidney transplant on a patient suffering from T2DM-induced end-stage renal disease) are likely to have lower impact and cost-effectiveness than a disease-specific analysis might suggest – with the lower impact reflecting the fact that, for example, such interventions typically require disease-specific treatment from specialized providers that offer little benefit for other diseases, and the fact that the benefits of such interventions are tempered by competing risks.

3 Aggregate Approaches

Within this paper, we distinguish population models according to their granularity of representation [29], classifying them either as *aggregate* or *individual*. Aggregate models stratify a population into one or more equivalence classes according to their attributes (state and parameter values) – a practice termed attribute-based disaggregation in [29]. Each equivalence class in such a model is treated as "well mixed", in that individuals within that class are not distinguished by the model. While it is possible that the resulting equivalence classes can be very small, most such models make use of a number of equivalence classes that is far smaller than the count of individuals within the population. For this reason, we term such models *aggregate* models.

Because of their smaller size, aggregate approaches have the distinct advantage of allowing for more rapid simulation. In addition to convenience, reduced simulation time also fosters improved understanding, by allowing for interactive exploration of models, more rapid checking of hypotheses. The smaller size of such models also improves understanding by simplifying observation of and reasoning about the model. Because of the smaller number of parameters involved, smaller model also generally afford easier parameterization and calibration (by reducing the risk of overfitting). Taken collectively, the time savings allowed by aggregate approaches can allow for greater refinement of model scope or formulation.

This section examines three aggregate representations for comorbidities, taking diabetes as an example. The first two representations characterize the independent state of each comorbidity without explicitly seeking to keep track of the number of individuals suffering from combinations of conditions. The third aggregate representation explicitly models individuals with these combination of disease conditions.

3.1 Approach 1: Aggregate Attributability

Within the first approach, we make use of an extremely high-level approach to quantifying the impact of our diabetes-specific intervention. Within this approach, we quantify the overall comorbidity burden associated with diabetes. Given that diabetes-specific interventions will not entirely eliminate this burden, we provide a rough proportional estimate of how much of that burden would be lifted by successful interventions on an individual.

Figure 1 shows an aggregate approach for representing diabetes comorbidities that explicitly models the progression of diabetes (dysglycemia) alone. Conditional probabilities of an individual being in a specific stage of each other comorbidity (here, normal kidney function, as measured, for example, through the estimated glomerular filtration rate) are used to relate diabetes progression to the progression of those comorbidities.

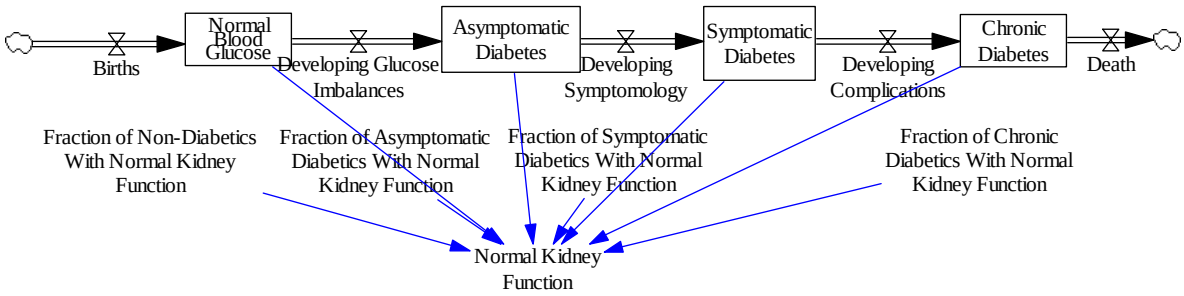

Figure 1: Aggregate Approach 1

While this approach is conceptually simple and maintains our modeling focus on diabetes, it suffers from several shortcomings.

- **No capacity for feedback.** Some comorbidities that are affected by diabetes also end up affecting diabetes. An example would be periodontal disease.
- **No comorbidity autonomy.** The representation provides no capacity to represent the natural progression of comorbidities (e.g. coronary artery disease, congestive heart failure) even in the absence of worsening of diabetes.
- **Inability to evaluate comorbidity-targeted interventions.** As a reflection of the above, the model will not be suitable for examining the impact of an intervention specifically targeting progression of a comorbidity (e.g. kidney transplant), even if it does nothing for diabetes.
- **No representation of individuals based on co-occurrence of conditions.** Depending on the tradeoffs involves, it may be desirable to focus an intervention on individuals at a certain stage of with respect to several conditions – for example, targeting advanced diabetics with mid-stage renal disease.

- **No Elimination of Key Heterogeneity Shortcoming.** One of the central motivations (mentioned above) for explicitly representing comorbidities was to allow us to better capture the impact of of heterogeneity on intervention effectiveness. Specifically, we sought a way to move away from the assumption that individuals with diabetes suffer from merely some general average level of co-morbidities. Here, we lack a way to capture the fact that the people we are helping do not have the "typical" patterns, but are, e.g. on average less likely to have nephropathy than your average diabetic at that stage – but more than a "normal person". For example, imagine that we could treat symptomatic diabetics with sufficient effectiveness that they become asymptomatic. While we might classify those successfully treated by the intervention as in the asymptomatic stage of diabetes progression, these affected individuals would not, in fact, be "average" asymptomatic diabetics – instead, they might be more likely to suffer from higher levels of renal disorders than would the average asymptomatic diabetic.

3.2 Approach 2: Coflows

Based on these concerns, we examine another representation, shown in Figure 2.

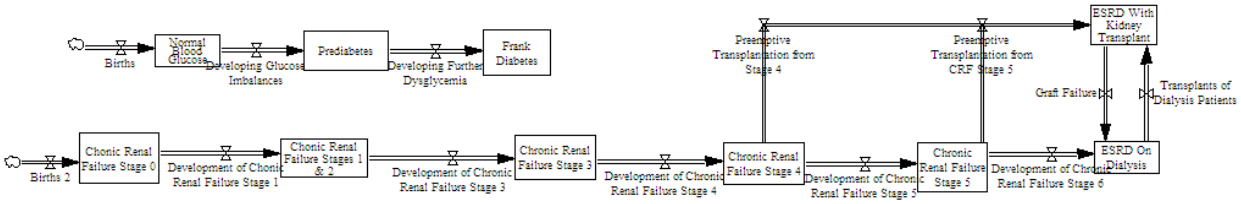

Figure 2: Co-Flow Representation.

The co-flow representation explicitly represents progression of each comorbidity. This representation enjoys some significant benefits over our initial attempt: Comorbidities can influence the progression of diabetes, and the other conditions can progress independent of diabetes progression. It is also possible to simulate the effect of interventions that directly affect the comorbidity (or example, to simulate the impact of adjusted dialysis regimen, or greater availability of kidney transplants.)

While the fragment appearing in Figure 2 looks suspiciously like a classic co-flow, it differs significantly. In a classic coflow, we are keeping track of the mean value of a characteristic for items in some stock. Here, we are instead trying to keep independent track of the distribution of the population in different stages of dysglycemia and nephropathy (whether caused by diabetes or of other etiology).

Unfortunately, this representation exhibits some significant shortcomings. Because there is no linkage in place to keep track of the number of individuals who are simultaneously in both a certain stage of diabetes and a certain stage of each of the other comorbidities, we can't target a particular class of individual (e.g. chronic diabetics with ESRD). Interventions targeting those with ESRD may be investigated, but only with the understanding that such interventions will affect those both with late stage diabetes and those without.

A related problem also arises from the latent heterogeneity of the stocks. A given person in the population will appear in some stock in each of the aging chains. When an intervention allows an individual to stay in the prediabetic stage of dysglycemia (rather than continuing on to frank Diabetes), we are again assuming that those affected by the interventions are like "average" symptomatic prediabetics with respect to chance of being at different stages of kidney disease, etc.)

A deeper issue here – and a common hurdle in creating an aggregate model from individual-level data – is that creating the appropriate probability distributions to be used in the model fragment in Figure 2 is quite challenging and labor intensive. It also does not seem fully natural – if we have access to individual-level data on people's conditions with respect to several conditions, it would seem more natural if we could just place them in an appropriate stock of the model that would capture that information, rather than treating them as spread between all stocks. The next section examines an approach – while still maintaining an aggregate perspective – that accomplishes just this.

3.3 Attribute-Based Disaggregation

We have seen from the above that while aggregate models have many virtues to recommend them, they do fall prey to some shortcomings. Specifically, such models can overestimate intervention impacts (by assuming that those that are affected by the intervention are "average" members of some population) and can require complex statistical relationships between the distributions of disease with respect to different conditions.

This subsection examines an alternative approach, one that explicitly captures the co-occurrence of comorbidities within some members of the populations via disaggregation. Within this approach, we stratify the population by attributes. Specifically, individuals are classified into equivalence classes according to static characteristics strongly impacting their level of risk (e.g. sex, ethnic group). Additional dynamic factors could also be taken into account (e.g. progression of diabetes, age, education level). The status of an individual with respect to each comorbidities could be captured by separating for each class of individuals those in different stages of coronary artery disease (or at least distinguish those with or without heart disease), stages of congestive heart failure, nephropathy, etc.

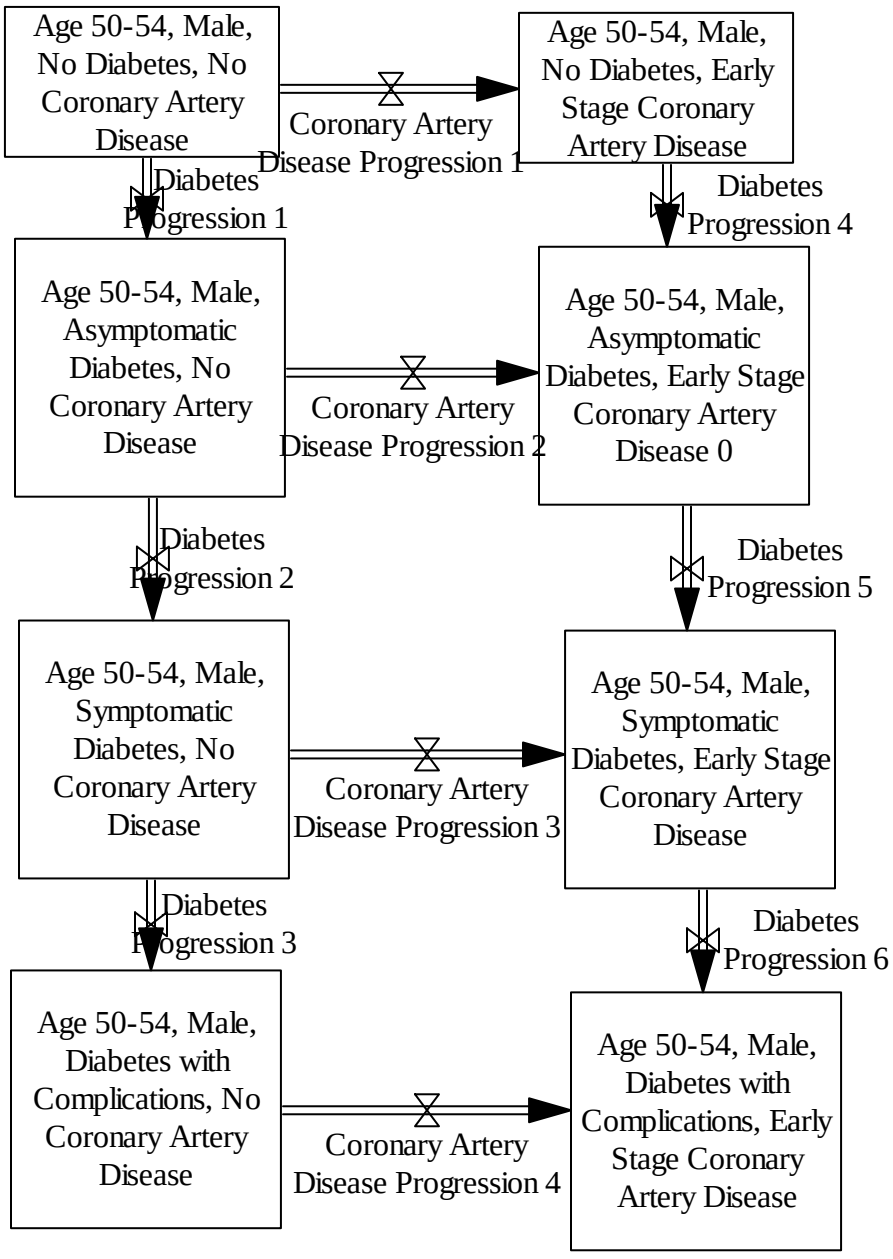

Figure 3: Attribute-Based Disaggregation

Figure 3 shows an example of an attribute-based disaggregation of diabetes and comorbidities. It can readily be appreciated that the number of stocks required is very large. If we made only binary distinctions for each just *n* categorization (including stages of diabetes), we would have 2^n stocks. Given

that individuals can leave a given stock by many means (e.g. aging, progressing to a later stage of any of the diseases), the number of flows between stocks would much greater yet.

While such a model can in principle represent complex causal interactions between stages of the disease – for example, capturing the fact that chronic diabetes will generally worsen heart and renal problems – the representations of these factors involves considerable mechanism. Such effects could be captured, for example, via a D-dimensional transition matrix (where D is the count of conditions of concern), but only at the cost of obscuring aspects of model structure through use of a somewhat opaque model formulation in which any stage of one disease can influence any stage of all other disease (simply with different weights).

While conveniently expressive, the arrayed set of stocks proposed would also obscure the important distinction between static attributes (parameters such as the sex or ethnic group of an individual) and dynamic attributes (such as age or stage of disease). All such attributes are used to determine the equivalence class in which an individual in the population must be counted. Progression would thus occur along several (dynamic) dimensions, but not along several others (those associated with static attributes). These progressions would be captured as different sorts of flows – for example, flows due to aging, and separate flows for progression of each disease, and potentially flows associated with behavior change (for example, cessation from or relapse into smoking). The detail complexity of the resulting structure would hinder understanding, analysis and visualization.

An additional concern relates to the modifiability of the representation. Consider adding a new attribute whose effect on disease risk or progression we wished to take into account – for example, levels of education or income. Adding such a parameter would require adding a new array dimension to the model. Adding this dimension would not only significantly increase the total number of states within the model but would also require re-specifying all of the associated flows between particular stocks.

3.4 Conclusion: Aggregate Approaches

The aggregate approach lends itself well to representation of a fairly homogenous set of individuals transitioning between well-defined discrete states. It tends to fare poorly when we need to distinguish between the parameter values and states along several different dimensions of heterogeneity. Representation of an individual flowing amongst continuous states using the aggregate approach is also somewhat awkward, as we must artificially discretize the process; the structure of our aggregate model may be shaped by the details of the transition process.

An aggregate-level model would also confront challenges in effectively modeling interpersonal factors. While it is relatively straightforward to capture at an aggregate level one-time vertical transmission of attitudes or dispositions occurring earlier in life (e.g. genetics, and in-utero exposure [16, 17], such as to gestational diabetes [21, 30-38]), the ongoing interpersonal interactions that play an important role in shaping risk factors will be captured with lower fidelity. Such factors are of central importance for communicable diseases, and for behavioral risk factors associated with many chronic diseases. While aggregate representations can be used, such representations typically require making some form of uniform mixing assumptions and do not allow for readily examining the impact of interventions aimed at particular points in social networks. Aggregate representation would, for example, be challenged to examine the ongoing impact of targeting at risk young women [39-41] on changing risk behavior of their children (specifically). Similarly, an aggregate model will exhibit difficulties I capturing the effects of

innovations in contact tracing methodology, or interventions at a particular point in a needle-sharing network (such as within a family of needle sharers, or for highly connected individuals in the intravenous drug use network). While data on individual-level characteristics is frequently available (e.g. via case-contact databases for notifiable illness, or from administrative or surveillance data), deriving an appropriate aggregate relationship that suitably captures the heterogeneity involved may require considerable reasoning.

Having briefly surveyed some of the challenges to aggregate-level modeling of such chronic disease issues, we turn now to examination of individual-based approaches.

4 Individual-Based Approaches

When dealing with simulation of larger populations, individual-based models are an important alternative to aggregate-level models. Such models are particularly attractive in the context of individually-targeted interventions, complex and dynamic networks linking individuals, high heterogeneity, and memoryful transition processes. Successes with applying multi-scale simulation models for infectious disease [42] suggests that the fine granularity of such models makes them particularly valuable for modeling the effects of peers, social norms, the physical environment, social networks and the community structure on individual behavior, and for simulating geographically situated populations. When considering multiple conditions in highly heterogeneous populations, such models frequently offer improved transparency, more graceful scaling in performance and space demands. As discussed in the previous Section, attributed-based aggregate-level models partition the population into equivalence classes distinguished by both parameters (for example, an individual's sex) and states (e.g. level of progression of a different diseases, exercise and nutritional habits). This leads to rapid proliferation of the set of stocks and flows required – and the resulting computational demands – as the number of parameters rise. An individually-disaggregated state space model (for example, that shown in Figure 4), can explicitly maintain separate information on parameters and states. An individual formulation has the virtue of explicitly separating out the static parameters from the dynamic attributes, yielding a cleaner representation of an individual's evolution over time. The individual-based model also opens the door to representing coupling between individuals (for example, the impact of a parent's risk factors on those of children or peers).

There are significant concerns involved in adoption of a multi-scale model, tradeoffs that have been in past [29] and ongoing work. Such models also offer greater flexibility – but all of the benefits come at a very heavy performance cost [29]. Despite the identification of techniques for reducing the performance burden through the development of explicit scale models [43], multi-scale models remain slower and more difficult to calibrate and analyze than their aggregate counterparts [44]. Because of the limited time available for modeling projects, the extra time required for model simulation, analysis and calibration imposes a high opportunity costs by taking time away from model refinement, and limit the expansion of model scope. As noted above, these shortcomings can have significant impact on the quality of one's insight into the model.

4.1 Individual-Based Approaches

There are many ways of achieving such an individual-based representation of comorbidities, each with their own significant tradeoffs. This paper presents three approaches. These approaches throw interesting light not only on chronic disease modeling concerns but also on methodological tradeoffs, particularly

between aggregate- and individual-level representations of populations and between distinct means of representing individual-level behavior. It is notable that all of these models have the advantage of being (primarily) declarative – that is, they allow the modeler to focus on the "what" of the model rather than on the details of "how" it is accomplished. We pause briefly to discuss this point.

There are two broad options for specifying individual-level models: Algorithmic techniques based around code (traditionally – but not always – written in an imperative, object-oriented language), and declarative mathematical techniques based around equations. Declarative, equation-based techniques offer some significant advantages over algorithmic specifications, including greater transparency, expressiveness, performance, scalability, modifiability, and ease of formal analysis, calibration, generalization, and formal verification [44]. We briefly review some key ideas here.

Declarative mechanisms allow a modeler – and model reviewers – to focus on the model being specified, rather than on the details of how it is to be solved. Declarative techniques also foster greater expressiveness and reduce the risk of many types of errors in model specification and modification. Finally, by making model logic more transparent to the associated executable engine (and compilation mechanisms) declarative mechanisms permit greater opportunities for performance optimization and scalability.

Explicit mathematical specification of a model permits formal reasoning about causes of a model's behavior – fostering improved analysis (e.g. deriving the location and stability of equilibria), ease of closed-form equilibration and calibration. A mathematically transparent model may also be more readily generalized, and permits frameworks to offer cross-checking mechanisms such as unit checks, interval analysis.

Within the next several subsections, we examine different individual-level approaches to representing comorbidities.

4.1.1 Approach 1: State Equations

Within the first approach to individual-level representation, we characterize the behavior of each individual using a set of parameters and state equations (differential equations). In contrast to traditional aggregate-level models (where individuals are classified into a set of discrete equivalence classes), the state of that individual is a point in a continuous state space, and the state evolves along some trajectory according to these state equations.

Within this approach, individuals can be coupled together through their state equations, in a classic form. Such inter-individual coupling will typically be far looser than the coupling that goes on between stocks within an individual. Generally speaking, it will be necessary to represent all such possible couplings using a matrix formulation, with the establishing and breaking (or attenuation) of connections implemented by changing the strength of connection between a state variables of one individual and the flows of another.

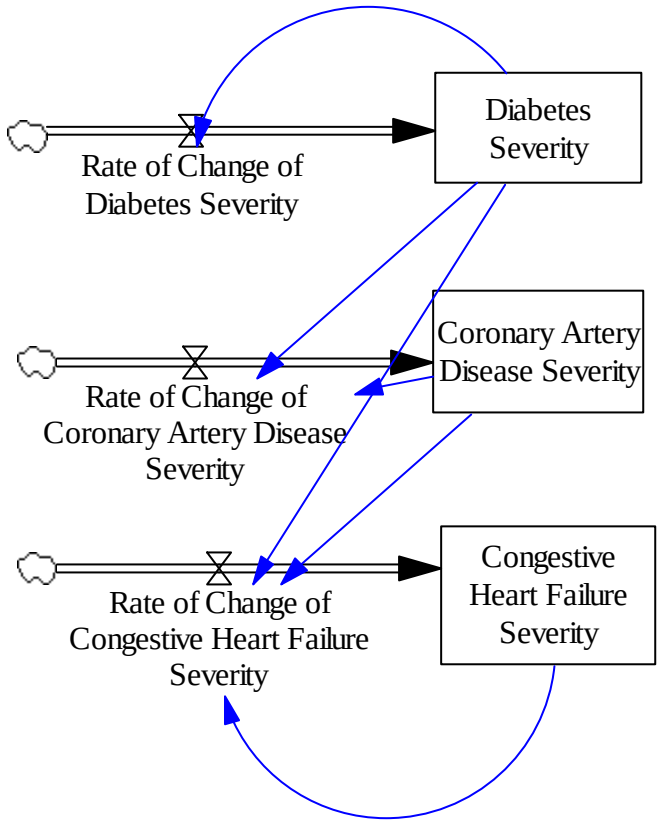

Figure 4: State Equation Representation of Continuous Disease Progression

Figure 4 shows how (at a schematic level) an individual-level model of comorbidities might be structured at the individual level. Within the context of the full system, the population will be composed of a collection of individuals. Each of these individuals is associated with explicit parameters representing static attributes (e.g. sex, ethnic group) as well as stocks for changing attributes such as age, progression of several conditions (measured on some unit-indexed scale). Another example of an individual-based system dynamics model (from the infectious disease literature) of this sort may be found in [42]; for this immunological models, the stocks represent immunological quantities of interest, and the flows of one stock (free virions) are coupled to the stocks of other people In the population.

The given model is a convenient representation of multi-scale effects involving both an individual and the broader population, and lends itself to sophisticated characterization of the coupling of multiple conditions.

4.1.2 Approach 2: Hybrid Automata

One shortcoming of the description of an individual's evolution using (individually-disaggregated) state equations is that it provides no convenient means of capturing qualitatively discrete state transitions. Such transitions might be associated with broader changes (regime switches or phase transitions) that modify the set of processes that we wish to take into account for an individual. For example, consider an individual who is born, undergoes puberty, who develops a complex disease, or who dies. These

transitions are associated with qualitative changes in the set of equations that govern behavior – much as the processes governing the evolution of a raindrop's shape change when the drop hits the ground.

In addition to such qualitative shifts, there are situations in which representing discrete states is convenient from the point of view of modeling simplicity. For example, in situations in which we have only a rough, qualitative notion of the stages of an illness, it is convenient to have recourse to a rough, discrete representation of disease progression. Such progression can be represented in a continuous time individually-disaggregated state equation framework, but only in a somewhat awkward fashion (such as through the use of binary states, or states whose values hold the probability associated with presence in that state).

Within this section, we describe *hybrid automata*, which combine together the continuous time and continuous state framework of state equations with discrete transitions and time. Like individually-disaggregated state equations, hybrid automata form a valuable alternative to aggregate models, but an alternative with greater flexibility than state equations.

Informally, a hybrid automaton consists of a countable set of states, and condition-driven transitions between those states. Within a given state, the system is governed by a specific set of differential equations. These differential equations will in general differ between different states; transitions between states can thus change the governing behavior of the states.

There are many alternative formulations of hybrid automata. We follow Henzinger [46] in defining a hybrid automaton as a 7-tuple (X, V,E, init, inv, flow, jump, events). The components of the hybrid automata are as follows:

- X: a set of real-valued state variables, visible both to other automata and within this one.
- V: a set of discrete states (control modes) for the automata, each associated with some transitions.
- E: a set of transitions ("control switches") between the control modes
- flow: $V \rightarrow \{dX/dt \rightarrow$ Equations involving state variables X$\}$. Specifies for each state variable the different equation governing its evolution within this control mode.
- Init: $2^{V \rightarrow \{(X,\Re)\}}$A set of initial conditions for the values of X. Upon entering a given control mode, a state variable may be initialized to some constant value.
- Inv: $\{ V \rightarrow$ Predicate involving x$\}$. The invariant specifies for every control mode a predicate that must remain true for the system to remain in that state.
- Jump conditions: $\{E \rightarrow$ Predicate involving x$\}$. Specifies the conditions under which a given control switch will be triggered. These conditions
- Output Events: $\{E \rightarrow \Sigma\}$. Labels each control switch with an output (issued) event.
- Input Events: $\{E \rightarrow \Sigma\}$. Labels each control switch with an input (received) event.

There is a large and well-developed theory of hybrid automata in control theory, as applied to hybrid digital and analogue systems. This theory has developed formalizations of interacting, concurrent automata, a rich set of analysis techniques, as well as techniques for formally simplifying automata. This technique deals with both deterministic and non-deterministic systems.

Separate hybrid automata generally interact by means of events and state variables. In some hybrid automata formulations, a distinction is made between those variables that externally vs. internally visible. (See, for example, [47]).

We believe that support for the full generality of hybrid automata is not required to deliver value in systems modeling. In particular, we believe that simplified, purely deterministic model that lacks invariants (but retains jump conditions and events would suffice). Such a formalism would allow for leveraging the great expertise that has evolved for modeling hybrid systems with minimal specification mechanism.

In summary, hybrid automata capture the need for modelers to use both discrete and continuous mechanisms, are declarative and easily understood, are subject to rigorous analysis, and represent a natural generalization of differential equations. Such automata can easily be augmented with dimensional information and other metadata. We believe that individual-level systems models could benefit considerably from wider use of such techniques.

Given that both hybrid automata and state equations (as realized through stock and flow diagrams) have rich visual represents, an interesting question relates to how to best combine these techniques for an intuitive visual understanding of a model.

4.1.3 Approach 3: Parallel State Transitions

A third approach to representation of cormorbidities uses (like the aggregate model) a progression of discrete states for each model.

Comparison between this approach and that of the aggregate model is instructive.

As we have seen above, within the aggregate model, the only way to maintain information on the fact that an individual has several attributes (e.g. specific stages of a number of comorbidities) is to count that individual as a member of some equivalence class of individuals who have specific ranges of values for each of those attribute. Such attribute-based aggregation becomes highly expensive, as the number of equivalence classes (and, thus, stocks) rises geometrically with the number of comorbidities (or other dimensions) considered [29].

Representing the discrete stages of comorbidity progression at an individual-level model is far less demanding. In contrast to the combinatorial explosion seen in the aggregate case, adding additional comorbidities requires just the addition of an additional state transition sequence. As a result, the resource needs only rise linearly with the count of comorbidities.

It is straightforward to represent the impact progression in a specific comorbidity has on other comorbidities. In the message-driven framework typical of agent-based modeling, this interaction can be realized in a “push” or “pull” fashion.

As a “push” framework, each stage of a specific comorbidity can lead to different likelihood that a ‘progress’ event will be sent to other comorbidities. For example, presence in a chronic stage of diabetes could lead to a relatively high probability density that heart disease would worsen. The probability density would be lower for earlier stages of diabetes.

As a "pull" framework, each transition between stages of disease could make use of a dynamic rate constant. The rate constant associated with a particular advance would depend on the stages obtaining for other comorbidities. To continue the example from earlier, the likelihood density of advancing to a later stage of heart disease could depend on the individual's current stages of diabetes and periodontal disease.

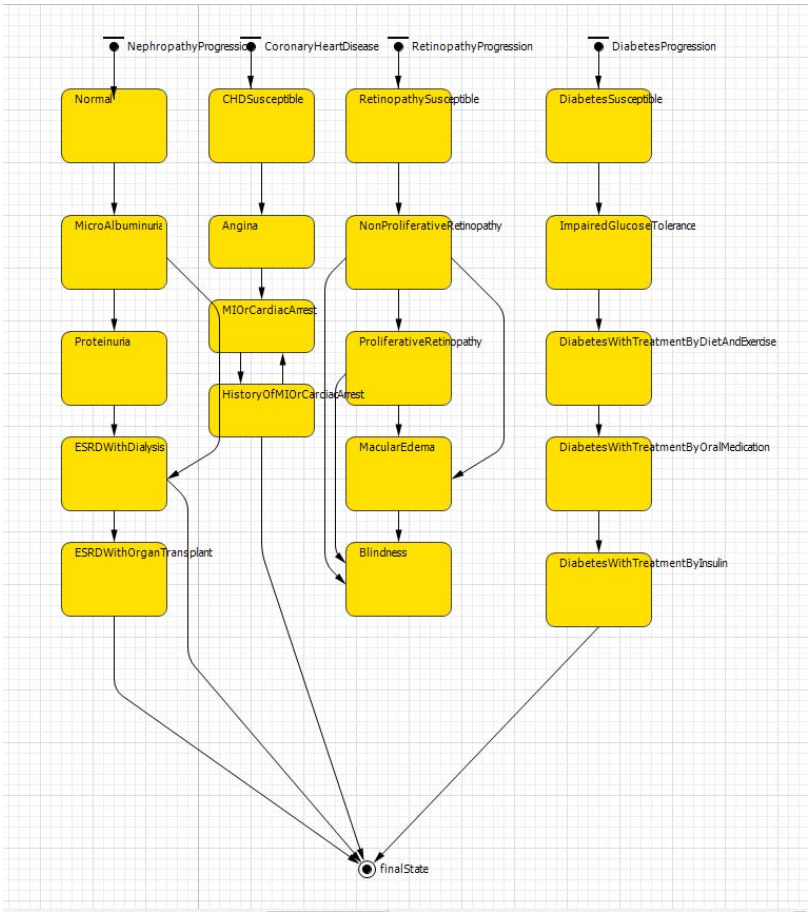

Figure 5: Representing Parallel Statecharts to Represent Co-Morbidities

The representation shown here allows for a transparent, orthogonal and modular decomposition of co-morbidities that avoids the combinatorial blowup associated with attribute-based disaggregation, while still allowing for straightforward capturing of co-morbidity interactions. The approach is not without its shortcomings. Reliance on discrete states and transitions inhibits symbolic reasoning about behavior, such as might be desirable to explain observed high-level dynamics, or to identify the location and stability of equilibria. Nonetheless, the representation provided here offers attractive simplicity as the dimensions of heterogeneity grow.

5 Capturing Inter-person Effects Using Multi-Scale Models

Within the section above, we have focused on the challenges associated with representing co-morbidities within a single individual in isolation. While the desire to capture high amounts of heterogeneity motivates the use of individual-level models representation of co-morbidities within individuals, such

models also offer considerable advantages when representing inter-personal effects. Within this section, we explore how individual-based model structure can ease capturing such effects.

One area of health in which inter-person impacts are central is communicable disease. Within recent years, there has been a pronounced growing trend towards the use of individual-based models in infectious disease epidemiology. This shift is motivated in large part by the desire to more precisely capture the effects of targeted intervention, the impact of the (typically pronounced) heterogeneity in contact rates, or dynamics of infection in early phases of an epidemic. While not yet well explored in the literature, co-morbidities present additional motivations for the use of individual-based models. For example, models representing co-infection or superinfection with multiple strains of Human Papilloma Virus or Herpes Simplex Virus can benefit from the orthogonal decomposition seen above. Other motivations lie in representing the impact of chronic disease on risk and natural history of infection.

While the effects of individual-individual and individual-community interactions on overall population health are less obvious for chronic disease than such effects are for infectious disease, they are still of critical importance. Individual behavior, diet, activity levels, and other factors influencing health are shaped social norms and expectations, affected by socially-influenced stress and self-esteem levels, modeled after peers, parents and communities, etc.

Explicitly representing social connections between individuals can also aid in reasoning about intervention effectiveness. For example, interventions targeting exercise or nutritional habits of young mothers may help their existing and future children. Interventions aimed at obese youth may help encourage health consciousness among their peers.

The causal interaction between two individuals, or within a larger group can often be elegantly captured within individual-based models with one or more networks. Within such models, different types of network edges pertinent to the spread of risk attitudes can be defined (e.g. those representing social interactions [42], networks representing family groupings, interactions between peers, etc.). Maintaining such network structures can permit the formulation of interventions that explicitly target particular members of a network (e.g. highly influential or contagious individuals), allows for tracing of the percolation of effects across the network. In the context of diabetes and its comorbidities, such effects allow us to reason about, for example the impact of improved nutrition or exercise on the part of parents on their children, or of peers' habits or weight on each other's behavior.

Such effects are difficult to capture in an aggregate format discussed above. Because each of the stocks is taken as "well-mixed", it is not possible to explicitly link the behavior of one individual to that of other individuals. Certain effects could be approximated statistically – for example, we may reason that causing a mother to improve her nutrition may, on average, cause behavior change in 0.2 of her children. But given that much data on individual behavior are formulated at the level of individuals, deriving such aggregate statistical relationships and putting into place the mechanisms to maintain these relationships often requires additional work and reasoning beyond that which is required for an individual-level model. (For example, ensuring the ongoing effects of a parent's changed behavior on a child may require rather complex linkages between flows associated with stocks at different ages. It also may require deriving the statistical relationship between a child's age and their parents' ages.) In cases where there are likely to be multiple and overlapping social structures that shape risk attitudes (e.g. family, community), it will be challenging to derive high-level relationships.

6 Conclusion

The need to represent complex interaction of conditions and interpersonal factors in heterogeneous populations is common to many health concerns. Both chronic disease and infectious diseases are associated with strong interactions between conditions that can have a first-order bearing on health. Failure to capture multiple comorbidities and their interactions within a model could lead to significantly overestimating or underestimating of the impact of interventions. Such concerns are of particular relevance for upstream interventions, which affect many downstream conditions – conditions that frequently co-occur.

Within this paper, we have explored several comorbidity representations. While individual representation exhibit particularly high resolution and transparent representation of co-morbidities, they have significant drawbacks that slow the modeling, and therefore frequently compromise model insight and refinement.

At the current point, the choice the most representation most suitable to a given problem is unclear. We believe that exploring a wide variety of representations (both aggregate and individual-based) within research, to discover research domains in which particular representations offer comparative advantage.

7 Bibliography

[1] L. L. Lipscombe and J. E. Hux, "Trends in diabetes prevalence, incidence, and mortality in Ontario, Canada 1995-2005: a population-based study," *The Lancet,* vol. 369, pp. 750-756, March 3 2007.

[2] M. Lemstra, C. Neudorf, and J. Opondo, "Health Disparity by Neighbourhood Income," *Canadian Journal of Public Health,* vol. 97, p. 5, November-December 2006.

[3] A. Boutayeb and A. Chetouani, "A critical review of mathematical models and data used in diabetology," *BioMedical Engineering OnLine,* vol. 5, p. 43, 2006.

[4] D. M. Eddy and L. Schlessinger, "Archimedes: A trial-validated model of diabetes," *Diabetes Care,* vol. 26, pp. 3093-3101, November 1 2003.

[5] L. Schlessinger and D. M. Eddy, "Archimedes: a new model for simulating health care systems--the," *Journal of biomedical informatics,* vol. 35, pp. 37-50, 2002.

[6] J. B. Homer and G. B. Hirsch, "System Dynamics Modeling for Public Health: Background and Opportunities," *American Journal of Public Health,* vol. 96, pp. 452-458, March 1 2006.

[7] J. Homer and B. Milstein, "Optimal decision making in a dynamic model of community health," in *37th Annual Hawaii International Conference on System Sciences*, 2004, p. 11.

[8] D. Rees, "System Dynamics Modelling as a Tool in Healthcare Planning," in *The 2006 International Conference of the System Dynamics Society* Nijmegen, The Netherlands: International System Dynamics Society, 2006.

[9] A. P. Jones, J. B. Homer, D. L. Murphy, J. D. K. Essien, B. Milstein, and D. A. Seville, "Understanding Diabetes Population Dynamics Through Simulation Modeling and Experimentation," *American Journal of Public Health,* vol. 96, pp. 488-494, March 1 2006.

[10] J. Homer, A. Jones, and D. Seville, "Diabetes System Model Reference Guide," Sustainability Institute, Hartland, Vermont 2006.

[11] E. Weir and L. Lipscombe, "Metabolic Syndrome: Waist Not, Want Not," *Canadian Medical Association Journal,* vol. 170, p. 2, April 27 2004.

[12] B. Balkau, R. J. Jarrett, K. Pyorala, and E. Eschwege, "Fasting blood glucose and risk of cardiovascular disease," *Diabetes Care,* vol. 22, pp. 1385-7, 1999.

[13] L. S. Ho, J. Gittelsohn, S. B. Harris, and E. Ford, "Development of an integrated diabetes prevention program with First Nations in Canada," *Health Promotion International,* vol. 21, pp. 88-97, June 1 2006.

[14] D. S. Shepard and R. J. Zeckhauser, "Long-term effects of interventions to improve survival in mixed populations," *Journal of Chronic Diseases,* vol. 33, pp. 413-33, 1980.

[15] S. L. Maddigan, D. H. Feeny, S. R. Majumdar, K. B. Farris, and J. A. Johnson, "Understanding the Determinants of Health for People With Type 2 Diabetes," *American Journal of Public Health,* vol. 96, pp. 1649-1655, September 1 2006.

[16] T. Tengs, N. Osgood, and L. Chen, "The Cost-Effectiveness of Intensive National School-Based Anti-Tobacco Education: Results from the Tobacco Policy Model," *Preventive Medicine,* vol. 33, pp. 558-70, 2001.

[17] T. Tengs, N. Osgood, and T. Lin, "Public health impact of changes in smoking behavior: results from the Tobacco Policy Model," *Medical Care,* vol. 39, pp. 1131-1141, October 2001.

[18] A. Makroglou, "Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview," *Applied Numerical Mathematics,* vol. 56, p. 559, 2006.

[19] J. Homer, G. Hirsch, M. Minniti, and M. Pierson, "Models for collaboration: how system dynamics helped a community organize cost-effective care for chronic illness," *System Dynamics Review,* vol. 20, pp. 199-222, 2004.

[20] A. J. Karter, M. R. Stevens, W. H. Herman, S. Ettner, D. G. Marrero, M. M. Safford, M. M. Engelgau, J. D. Curb, and A. F. Brown, "Out-of-Pocket Costs and Diabetes Preventive Services:

The Translating Research Into Action for Diabetes (TRIAD) study," *Diabetes Care,* vol. 26, pp. 2294-2299, August 1 2003.

[21] R. F. Dyck, "Tracking ancient pathways to a modern epidemic: diabetic end-stage renal disease in Saskatchewan aboriginal people," *Kidney International,* 2005.

[22] T. K. Young, P. J. Martens, S. P. Taback, E. A. C. Sellers, H. J. Dean, M. Cheang, and B. Flett, "Type 2 Diabetes Mellitus in Children: Prenatal and Early Infancy Risk Factors Among Native Canadians," *Archives of Pediatrics and Adolescent Medicine,* vol. 156, pp. 651-655, July 1 2002.

[23] M. L. Brandeau, "Modeling complex medical decision problems with the Archimedes model," *Annals of Internal Medicine,* vol. 143, pp. 303-4, 2005.

[24] D. M. Eddy, L. Schlessinger, and R. Kahn, "Clinical Outcomes and Cost-Effectiveness of Strategies for Managing People at High Risk for Diabetes," *Annals of Internal Medicine,* vol. 143, pp. 251-264, August 16 2005.

[25] A. J. Palmer, S. Roze, W. J. Valentine, M. E. Minshall, V. Foos, F. M. Lurati, M. Lammert, and G. A. Spinas, "The CORE Diabetes Model: Projecting Long-term Clinical Outcomes, Costs and Cost-effectiveness of Interventions in Diabetes Mellitus (Types 1 and 2) to Support Clinical and Reimbursement Decision-making," *Current Medical Research and Opinion,* vol. 20, pp. 5-26, 2004.

[26] H. Zhou, D. J. M. Isaman, S. Messinger, M. B. Brown, R. Klein, M. Brandle, and W. H. Herman, "A Computer Simulation Model of Diabetes Progression, Quality of Life, and Cost," *Diabetes Care,* vol. 28, pp. 2856-2863, December 1 2005.

[27] D. M. Eddy and L. Schlessinger, "Validation of the Archimedes Diabetes Model," *Diabetes Care,* vol. 26, pp. 3102-3110, November 1 2003.

[28] American Diabetes Association Consensus Panel, "Guidelines for Computer Modeling of Diabetes and Its Complications," *Diabetes Care,* vol. 27, pp. 2262-2265, September 1 2004.

[29] N. Osgood, "Representing Heterogeneity in Complex Feedback System Modeling: Computational Resource and Error Scaling," in *Presented at the 22nd International Conference of the System Dynamics Society* Oxford, UK: International System Dynamics Society, 2004.

[30] R. Dyck, H. Klomp, L. K. Tan, R. W. Turnell, and M. A. Boctor, "A Comparison of Rates, Risk Factors, and Outcomes of Gestational Diabetes Between Aboriginal and Non-Aboriginal Women in the Saskatoon Health District," *Diabetes Care,* vol. 25, pp. 487-493, March 1 2002.

[31] R. F. Dyck, "Differences in High Birthweight Rates between Northern and Southern Saskatchewan: Implications for Aboriginal Peoples," *Change,* vol. 1977, p. 4, 1995.

[32] R. F. Dyck, L. Tan, and V. H. Hoeppner, "Body Mass Index, Gestational Diabetes and Diabetes Mellitus in Three Northern Saskatchewan Aboriginal Communities," *Chronic Diseases in Canada,* vol. 16, p. 3, 1995.

[33] L.-S. Fetita, E. Sobngwi, P. Serradas, F. Calvo, and J.-F. Gautier, "Consequences of Fetal Exposure to Maternal Diabetes in Offspring," *Journal of Clinical Endocrinology & Metabolism,* vol. 91, pp. 3718-3724, October 1 2006.

[34] P. W. Franks, H. C. Looker, S. Kobes, L. Touger, P. A. Tataranni, R. L. Hanson, and W. C. Knowler, "Gestational Glucose Tolerance and Risk of Type 2 Diabetes in Young Pima Indian Offspring," *Diabetes,* vol. 55, pp. 460-465, February 1 2006.

[35] S. Rodrigues, E. Robinson, and K. Gray-Donald, "Prevalence of gestational diabetes mellitus among James Bay Cree women in northern Quebec," *Canadian Medical Association Journal,* vol. 160, pp. 1293-1297, May 4 1999.

[36] S. Rodrigues, E. J. Robinson, M. S. Kramer, and K. Gray-Donald, "High Rates of Infant Macrosomia: A Comparison of a Canadian Native and a Non-Native Population," *J. Nutr.,* vol. 130, pp. 806-812, April 1 2000.

[37] B. R. Vohr, L. P. Lipsitt, and W. Oh, "Somatic growth of children of diabetic mothers with reference to birth," *The Journal of Pediatrics,* vol. 97, pp. 196-9, 1980.
[38] R. C. Whitaker, M. S. Pepe, K. D. Seidel, J. A. Wright, and R. H. Knopp, "Gestational Diabetes and the Risk of Offspring Obesity," *Pediatrics,* vol. 101, February 1 1998.
[39] H. Klomp, R. F. Dyck, and S. Sheppard, "Description and Evaluation of a Prenatal Exercise Program for Urban Aboriginal Women," *Canadian Journal of Diabetes,* vol. 27, p. 8, 2003.
[40] R. F. Dyck, M. Sheppard, H. Cassidy, K. Chad, L. Tan, and S. Van Vliet, "Preventing NIDDM among aboriginal people: Is exercise the answer? Description of a pilot project using exercise to prevent gestational diabetes," *International Journal of Circumpolar Health,* vol. 57, 1998.
[41] R. F. Dyck, M. S. Sheppard, H. Klomp, L. K. Tan, K. Chad, S. H. Van Vliet, and P. Paterson, "Using Exercise to Prevent Gestational Diabetes Aboriginal Women -- Hypothesis and Results of a Pilot/Feasibility Project in Saskatchewan," *Canadian Journal of Diabetes Care,* 1999.
[42] A. J. Rhee, "An agent-based approach to HIV/AIDS epidemic modeling : a case study of Papua New Guinea," Dual S.M.,Technology and Policy Program and Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Mass., June 2006. 133 p
[43] N. Osgood, "Enriching Dimensionality Analysis in System Dynamics Modeling," in *System Dynamics Winter Conference 2007*, Austin, Texas, 2007.
[44] N. Osgood, "Systems Dynamics and Agent-Based Approaches: Clarifying the Terminology and Tradeoffs," in *First International Congress of Business Dynamics* Brasília, Brazil, 2006.
[45] M. A. Nowak and R. M. May, *Virus dynamics : mathematical principles of immunology and virology*. Oxford ; New York: Oxford University Press, 2000.
[46] T. A. Henzinger, "The Theory of Hybrid Automata," in *11th Annual IEEE Symposium on Logic in Computer Science (LICS'96)*, 1996, p. 278.
[47] N. Lynch, R. Segala, and F. Vaandrager, "Hybrid I/O Automata," *Information and Computation* vol. 185, pp. 105-157 25 August 2003.
[48] XJ Technologies,Software Program "AnyLogic",ver. 6,St. Petersburg, Russia,2007
[49] J. T. Coffey, M. Brandle, H. Zhou, D. Marriott, R. Burke, B. P. Tabaei, M. M. Engelgau, R. M. Kaplan, and W. H. Herman, "Valuing Health-Related Quality of Life in Diabetes," *Diabetes Care,* vol. 25, pp. 2238-2243, December 1 2002.
[50] D. G. Manuel and S. E. Schultz, "Health-Related Quality of Life and Health-Adjusted Life Expectancy of People With Diabetes in Ontario, Canada, 1996-1997," *Diabetes Care,* vol. 27, pp. 407-414, February 1 2004.
[51] H. Klomp, B. Chan, P. Cascagnette, G. Teare, and N. Sidhu, *Quality of diabetes management in Saskatchewan: Supplementary tables and figures*. Saskatoon: Health Quality Council, 2006.
[52] D. C. Benyshek, J. F. Martin, and C. S. Johnston, "A reconsideration of the origins of the type 2 diabetes epidemic among Native Americans and the implications for intervention policy," *Medical Anthropology,* vol. 20, pp. 25-64, 2001.
[53] T. K. Young, J. Reading, B. Elias, and J. D. O'Neil, "Type 2 diabetes mellitus in Canada's First Nations: status of an epidemic in progress," *Canadian Medical Association Journal,* vol. 163, pp. 561-566, September 5 2000.
[54] Indigenous Peoples' Health Research Centre, "Miyo-Māhcihowin A report on Indigenous Health in Saskatchewan," Indigenous Peoples' Health Research Centre 2006.
[55] A. J. G. Hanley, S. B. Harris, M. Mamakeesick, K. Goodwin, E. Fiddler, R. A. Hegele, J. D. Spence, A. A. House, E. Brown, B. Schoales, J. R. McLaughlin, R. Klein, and B. Zinman, "Complications of Type 2 Diabetes Among Aboriginal Canadians: Prevalence and associated risk factors," *Diabetes Care,* vol. 28, pp. 2054-2057, August 1 2005.

[56] T. K. Young, B. Elias, J. Reading, J. D. O'Neil, A. Leader, and G. McDonald, "Chronic Diseases: Literature Review and Analysis of the First Nations and Inuit Regional Health Survey National Core Data," Centre for Aboriginal Health Research 1998.

[57] R. F. Dyck and L. Tan, "Rates and outcomes of diabetic end-stage renal disease among registered native people in Saskatchewan," *Canadian Medical Association Journal,* vol. 150, p. 6, January 15 1994.

[58] First Nations Centre, "First Nations Regional Longitudinal Health Survey 2002/2003: Results for Adults, Youth and Children Living in First Nations Communities," Ottawa, ON: First Nations Centre, 2005.

[59] M. P. Pioro, R. F. Dyck, and D. C. Gillis, "Diabetes prevalence rates among First Nations adults on Saskatchewan in 1990 - Comparison by Tribal Grouping, Geography and with non-First Nations people," *Canadian Journal of Public Health,* vol. 87, pp. 325-8, 1996.

[60] L. Potvin, M. Cargo, A. M. McComber, T. Delormier, and A. C. Macaulay, "Implementing participatory intervention and research in communities: lessons from the Kahnawake Schools Diabetes Prevention Project in Canada," *Social Science and Medicine,* vol. 56, pp. 1295-1305, 2003.

[61] B. Jeffery, S. Abonyi, R. Labonte, and K. Duncan, "Engaging Numbers: Developing Health Indicators that Matter for First Nations and Inuit People," *Journal of Aboriginal Health,* vol. 3, p. 44, September 2006.

[62] Y. Iwasaki and J. Bartlett, "Stress-coping Among Aboriginal Individuals with Diabetes in an Urban Canadian City," *Journal of Aboriginal Health,* vol. 3, p. 11, September 2006.

[63] Lancet, "Targeting high-risk populations in the fight against diabetes," *The Lancet,* vol. 369, p. 716, March 3 2007.