Challenges in Formulating Explanatory Models for Co-morbidities

Matthew Klenk and **Marzieh Nabi** and **Adam Arvay**
Johan de Kleer and **Daniel G. Bobrow**
Palo Alto Research Center
3333 Coyote Hill Rd
Palo Alto, CA, 94304
klenk,marziehn,arvay,dekleer,bobrow@parc.com

Abstract

Patients with multiple health conditions pose significant challenges for modern healthcare. Understanding if and how these conditions are linked is essential to providing effective treatment. Physicians and researchers create explanatory models to develop hypotheses for these connections. In this paper, we discuss the breadth of domains these explanations draw upon as well as the diversity of applications of these models. Throughout the paper, we use example explanatory models from published literature and discuss the state-of-the-art of knowledge representation to support clinicians.

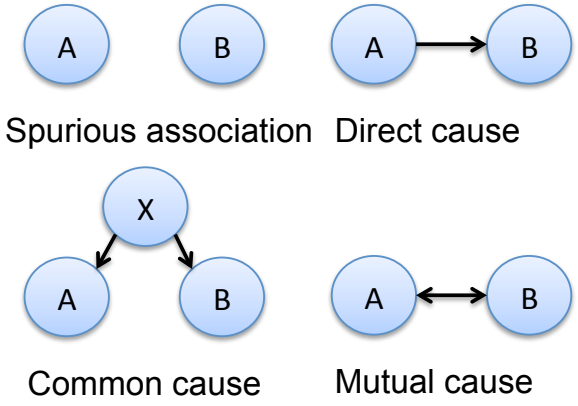
1 Introduction

Patients with multiple health conditions, or co-morbidities, pose significant challenges for modern healthcare. These patients make up 71% of total healthcare spending in America and 93% of Medicare spending.[1] To effectively treat patients with multiple health conditions, healthcare providers must consider if there is a causal relationship between the conditions and, if so, what mechanism underlies this relationship. Consider the hypotheses that obesity causes type 2 diabetes. Some treatments for type 2 diabetes cause weight gain (e.g., thiazolidinediones) potentially leading a vicious cycle of increasing insulin resistance (Kenkre, Tan, and Bloom 2013). To support healthcare providers, comorbidity researchers use electronic medical records to identify statistical relationships between diseases. Given a statistical relationship, it is necessary to understand if it occurs by either chance/sampling bias or through a causal mechanism. Figure 1 shows some of the different ways in which diseases may be causally connected.

Simply establishing a causal connection is insufficient. Effective treatment requires models of the underlying conditions and their interactions. Consider the hypothesized relationship between obesity and diabetes shown in Figure 2 (Liebman 2010). Under this model, increased levels of cortisol are the result

[1] `http://www.hhs.gov/ash/initiatives/mcc/final-whcoa-mcc-slides-remediated.pdf`

Figure 1: Possible causal relationships between diseases

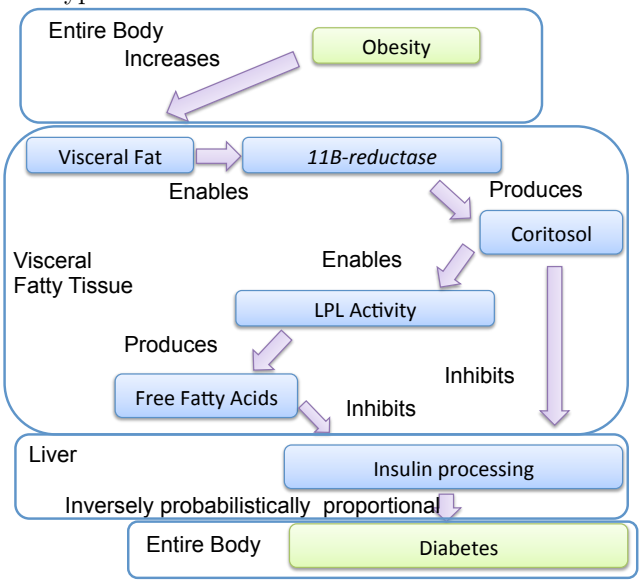


of increased 11B-reductase activity in visceral fatty tissue. The increased cortisol is transported from the fatty tissue to the liver through the portal vein. In the liver, cortisol promotes insulin resistance both directly and through the production of additional free fatty acids. Liver insulin resistance may spread throughout the body resulting in the patient having type 2 diabetes. Under this model, surgical treatments that remove visceral fat and drugs that regulate 11B-reductase activity or the level of cortisol will be effective treatments to break the link, but treatments that regulate the free fatty acid level in the liver will be insufficient.

While this explanatory model is an example of direct causation in Figure 1, when combined with the model of diabetes medicine causing weigh gain, the relationship is one of mutual causation. Simply identifying relationships between diseases with arrows and perhaps weights misses significant opportunities to support healthcare providers. We argue that representations of the explanatory models used by healthcare workers would enable new tools that could improve health outcomes.

Automatically constructing models to support these inferences is an exciting problem. The qualitative reasoning (QR) community's focus on understanding

Figure 2: Possible causal mechanism between obesity and type 2 diabetes.



the modeling process places it in a unique position to address this problem. QR researchers avoid ad hoc modeling in favor of reusable compositional models (Falkenhainer and Forbus 1991). The explicit representation of views enables component models referring to different levels of abstraction to be included in a single model. Furthermore, QR has developed rich models of causal reasoning that capture rules of thumb, probabilistic associations and well-understood mathematical relationships (Forbus 1984)(Weld and de Kleer 1989)(Kuipers 1994). While these features are essential for formulating the explanatory models identified in this paper, automatically creating them will require broadening established qualitative reasoning theories. The results of such an endeavor would not only transform the model-based reasoning community but also have significant impacts on medical research and practice.

This paper analyzes explanatory models found in literature to identify their properties and articulate the challenges. We make no claim as to the validity of these models, but include them to understand how practioners reason and communicate about comorbidities.

2 Example Comorbidity Explanations

In addition to the obesity and diabetes relationship described in Figure 2, we present three more explanatory models of comorbidities that will be referenced throughout the rest of this paper.

2.1 Autism and Asthma

Autism is frequently diagnosed before asthma. This could lead one to consider a direct causal relationship from autism to asthma. On the contrary, one proposed causal mechanism indicates that there might be common cause relationship between the conditions (Gidaya et al. 2016). In particular, Beta-2-adrenergic receptor inhalers (B2AR) are treatments for asthma. The use of B2AR during pregnancy has been associated with increased risk of autism developing in the child. Furthermore, asthma has a hereditary component resulting in an increased likelihood of children born to mothers using B2AR during pregnancy developing asthma and autism.

2.2 Diabetes and Lower Leg Amputation

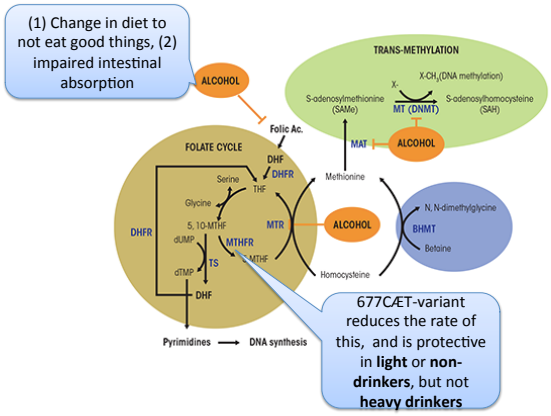
Given correlation between diabetes and lower leg amputation, Mayfield *et al.* (Mayfield et al. 1998) explore the likelihood of different mechanisms and how they affect treatment decisions. Here, we discuss a subset of the potential explanations involving the altered biomechanics of the patient. The following alterations in biomechanics can lead to ulcers and other lower leg trauma for which amputation is a treatment:

- Diabetics have increased body mass putting additional strain on the lower extremities.
- Lower leg trauma caused by limited joint mobility resulting from bone deformities and soft tissue damage.
  - Diabetes leads to bone deformities via motor neuropathy, the failure of neurons to communicate with certain muscles.
  - Diabetics have changes in their skin due to glycosylation, a metabolic process affected by diabetes, in soft tissue cells. These changes result in less pliable skin that is more prone to breaking.
- Neuropathy, a common symptom of diabetes, may prevent people from changing their gait as damage accumulates. Damage may be identified by temperature increase that would be perceptible to a clinician.
  - Peripheral vascular disease, which is frequently associated with diabetes, may lower foot temperature.
  - Neuropathy, a symptom of diabetes, may raise foot temperature.

2.3 Alcoholism and Cancer

Boffeta and Hashibe discuss the causal associations between alcohol consumption and different kinds of cancers (Boffetta and Hashibe 2006). While the mechanism is not well understood, the authors present two possible mechanisms acting through different metabolic pathways: ethanol and folate. Ethanol metabolism occurs in two steps. First, ethanol is transformed into acetaldehyde at a rate governed by the ADH and CYP2E1 gene families. Next, acetaldehyde is transformed into acetate at a rate governed by the ALDH gene. Genetic variation in ALDH gene

Figure 3: The possible effects of alcohol on folate metabolism.



family affect the rate of ethanol reactions by up to 90x. Alcoholism leads to increased alcohol consumption which in conjunction with genetic variation can lead to increased concentrations of acetaldehyde which is a known carcinogen. For example, the ALDH2 gene significantly slows the production of acetate allowing acetaldehyde to remain in the body at higher concentrations. Furthermore, this gene has been associated with increased risk for oral and throat cancers likely due to high concentrations of alcohol at those locations.

Figure 3 illustrates the possible effects of alcohol on folate metabolism. An important function of the folate cycle is DNA synthesis. Damaging this process increases the risk of developing cancer. First, alcohol reduces the amount of folate that enters the body. This is either through the poor diet of heavy drinkers or through alcohol affecting the intestinal absorption of folate. Within the folate cycle, Alcohol inhibits the expression of the MTR gene (Platek et al. 2009). The C677T variant of the MTHFR gene reduces the rate of 5-methylenetetrahydrofolate and appears to mediate the risk of colon cancer for light and moderate drinkers, but not for heavy drinkers.

3 Domains of Knowledge

From these example explanatory models, we identified six different domains of knowledge represented: genetic and metabolic pathways, physiology, mechanical, spatial, disease, and drug. For each domain of the knowledge, we present snippets from our examples and identify existing applicable biomedical knowledge bases.

3.1 Genetic and Metabolic Pathways

Genetic and metabolic pathways describe the chemical reactions that underlie biological phenomena. 11B-reductase from the obesity causing diabetes mechanism is a metabolic reaction. Proteins are the participants of metabolic reactions and genetic variation affects the rate which these proteins are transformed. Two different metabolic pathways, collections of reactions, have been identified as the possible mechanism concerning how alcoholism causes cancer.

Biologists are encoding the knowledge necessary to represent and reason about these pathways in wikipathways[2] and reactome (Joshi-Tope et al. 2005). Given the ontological structure of these models involving processes, rates, and concentrations, it is not surprising that members of the qualitative reasoning community have built systems to reason with this knowledge (Bredeweg et al. 2012). There is also work that links diseases to pathways that bioinformatics researchers have used to explain comorbidities through associations with the same pathways (Li and Agarwal 2009).

3.2 Physiological Models

Physiological models refer to the physical structure of the body and their functions. For example, physiological models are used multiple times in the associations between alcohol and cancer. First, when discussing the ethanol metabolic pathway, the explanation focuses on changes in oral and throat cancer rates due to the their roles in ingesting alcohol. Second, when analyzing reasons for decreased folate intake, the explanation discusses decreased intestinal function.

The majority of physiological modeling efforts have focused on linking genes and proteins to physiological functions (e.g., GO_MF (Ashburner et al. 2000) and Chemical Entities of Biological Interest (Degtyarenko et al. 2008)). There are ongoing efforts to link these ontologies to tissue-level descriptions (De Bono et al. 2015).

3.3 Mechanical Models

Mechanical models produce inferences from the physical connections of the body. That is, how the body moves and how different parts respond to forces applied to them. For example, the causal model concerning limited joint mobility in diabetics contains both static and dynamic models. Dynamic models include the fact that bone deformities restrict the range of movement of joints. Static models include the fact that changes in skin due to glycosylation increase the likelihood of breaking.

At this point, we are not aware of any reusable mechanical models of the human body that are used for healthcare. Standard practice appears to be to create a mechanical model for a specific purpose (Fung 2013). For reusable models, the most closely related efforts come from safety engineering (e.g., simulation of vehicle crashes) (Vezin and Verriest 2005).

[2] www.wikipathways.org

3.4 Spatial Models

Spatial models localize phenomena and interactions. While exact spatial locations may not be necessary, the representation of containers and connections is important. In diabetes model, the connection between visceral fatty tissue and liver through the portal vein is a central part of the explanation. Furthermore, the concentrations of different proteins must be understood with respect to a container.

The Open Biology Ontologies (OBO) include relationships for containment and adjacency for all of their ontologies (Smith et al. 2007). The Biological Spatial Ontology offers extensions to define precise, relative positions within an organism (Dahdul et al. 2014).

3.5 Disease Models

These models define diseases in terms of their signs, symptoms, and transmission. Disease symptoms can then be used to create patient specific models to identify how diseases may relate to one another. For example, in the obesity causes diabetes model, it is necessary to consider that obese people have more visceral fat. Then, the model is completed with the fact that increased insulin resistance is the defining signature of type 2 diabetes. Other examples of symptoms used in our examples include the fact that alcoholism has a symptom of increased alcohol consumption and reduces the amount folate in the diet. Representing disease transmission is necessary to form the causal link between the asthmatic mother and risk of asthma in their child.

The Disease Ontology is an ontology for describing the classification of human diseases organized by etiology, or causation (Kibbe et al. 2015). Alternatively, International Statistical Classification of Diseases and Related Health Problems (ICD-10) is used by many electronic medical record systems and contains codes for diseases, signs and symptoms, and abnormalities (Organization, Organization, and others 1992).

3.6 Drug models

Drug models describe how drugs interact with the body in multiple ways. Drug models that describe how a drug affects metabolic pathways may be used to design treatments. For example, a drug regulates 11B-reductase activity can be used to mitigate the risk of diabetes resulting from obesity. At the level of medical conditions and symptoms, drug models describe what diseases or symptoms drugs are used for and what their side effects are. For example, B2AR is a drug that treats asthma and that this drug has a side effect in pregnant women that increases the risk of autism in their children.

The National Institute of Health keeps records of drug interactions and side effects and makes this data available through APIs.[3] The OBO ontologies include multiple efforts to represent drugs and their effects. The DRON ontology supports comparative effectiveness researchers studying claims data.[4] The DINTO ontology categorizes drug-drug interactions (DDIs). This includes a representation of the possible mechanisms that can lead to them (including both pharmacodynamic and pharmacokinetic DDI mechanisms) (Herrero-Zazo et al. 2015).

In this section, we identified five different domains of knowledge that appear in our example explanatory models. In the next section, we highlight how this knowledge enables different types of inferences for healthcare providers.

4 Purposes of Explanatory Models

Explanatory models of comorbidities are important because they guide treatment decisions. Simple models of the form shown in Figure 1 in which nodes represent conditions and arrows represent causal relationships are insufficient for planning treatments. To illustrate, consider the abstract case of disease A causing disease B. If a patient has both diseases, will simply treating disease A be sufficient? Perhaps, but without an explanatory model to guide treatment this causal connection is not useful. In the rest of this section, we describe how explanatory models support model validation, patient observation, and treatment using examples.

4.1 Support for Model Validation

Medical researchers begin with statistical relationships between conditions. With the introduction of electronic medical records, bioinformatics researchers have developed new tools to identify orders of magnitude more potential relationships between diseases (Li and Agarwal 2009). In the diabetes lower leg amputation example, to validate the causal link between the increased body mass from diabetes and lower leg amputation, researchers conducted experiments to measure the peak plantar pressure in diabetic people. They found that body weight only accounts for less than 14% of the variance thus weakening the importance of this causal connection. This model validation step is essential in determining treatment decisions.

4.2 Support for Clinical Decisions

Healthcare providers use explanatory models to determine patient treatment and guide the monitoring of a single condition to ensure that it does not cause other conditions. In the asthma-autism connection, this involves changing the asthma medication for potential mothers. In the diabetes-amputation connection, the explanatory model explores if lower-leg stress can be identified through changes in temperature. The purpose of this statement is to support clinicians who are monitoring diabetic patients to identify those that

[3] `https://wwwcf2.nlm.nih.gov/nlm_eresources/eresources/search_database.cfm`

[4] `http://www.obofoundry.org/ontology/dron.html`

are at risk of developing ulcers that would lead to lower leg amputation.

Given a comorbid patient, explanatory models guide the treatment process. Instead of treating each condition in isolation, it is important to identify potential interactions between them. In the obesity-diabetes case, it is important that the treatment of diabetes does not lead to weight gain as that will counteract whatever treatment is being given to obesity.

5 So What? Advanced Tools for Clinical Support

In this work, we have identified a diverse set of knowledge domains necessary to create useful explanatory models of comorbidities. In Section 3, we illustrated some of the current efforts for creating reusable model libraries in each domain. From this landscape, we see two exciting research questions:

1. What kinds of inferences are possible from these explanatory models?
2. How can explanatory models be automatically constructed?

In the previous sections, we have already discussed aspects of the first question. A subset of the inferences that can be drawn from these models includes determining intervention decisions (e.g., not recommending B2AR inhalers for pregnant mothers), guiding future experiment design and data collection (e.g., determining the strength of the causal relationship between increased body mass and lower leg amputation), and directing healthcare monitoring (e.g., importance of watching weight for people taking diabetes drugs). Further research must explore the context of these decisions and other clinical decisions made by providers.

The second question concerns model formulation (Falkenhainer and Rajamoney 1988)(Rickel and Porter 1997). Given a question, this process typically involves (1) generating a model from a domain theory and experience, (2) evaluating its utility, and (3) revising the model based on its evaluation. Steps 2 and 3 continue until the modeler is satisfied with the results. One area where current work falls short for our application concerns the representation of the evaluation criteria and how these can change during the model revision process. Instead, current approaches typically address prediction questions (e.g., "What will happen to a quantity in a particular scenario?"). The explanatory models described in this work are often exploratory in nature (e.g., "How are these two conditions related?"). Their construction is important for communication between scientists, providers, their patients, and the public.

Understanding the context and inferences that are important to healthcare providers coupled with new techniques of model formulation and revision could enable new classes of clinical support tools.

6 Discussion

In this work, we argue that the simple causal network models that are shown in Figure 1 are insufficient for clinical support. Scientists and healthcare providers create explanatory models that expand the thin arrows in Figure 1 into mechanistic explanation consisting of entities and relationships.

The qualitative reasoning (QR) community's focus on understanding the modeling process places it in a unique position to bridge the gap between the causal network models that are derivable from data and the explanatory models used by clinicians. Compositional modeling's emphasis on reusable components, or model fragments, (e.g., 11B-reductase reaction is studied in isolation from diabetes patients) and explicit representation of assumptions are essential components of explanatory models. QR has developed rich models of causal reasoning that capture world knowledge (e.g., treatments of a pregnant mother affect the fetus), probabilistic associations (e.g., increasing insulin resistance increases the likelihood of a patient exhibiting diabetes) and understood mathematical relationships (e.g., systems biology models of chemical reactions, such as 11B-reductase, use differential equations).

While these features are essential for formulating the explanatory models identified in this paper, we also illustrate additional challenges that will require extending current QR theories and research. In particular, the extension of automated model formulation and revision from prediction and system identification tasks to the open-ended problem of comorbidity explanation. Such advances could enable a radical transformation of clinical support tools signficantly improving healthcare outcomes.

Acknowledgments

This work is supported by the Xerox Innovation Group.

References

[Ashburner et al. 2000] Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. 2000. Gene ontology: tool for the unification of biology. *Nature genetics* 25(1):25–29.

[Boffetta and Hashibe 2006] Boffetta, P., and Hashibe, M. 2006. Alcohol and cancer. *The Lancet Oncology* 7:149–156.

[Bredeweg et al. 2012] Bredeweg, B.; Linnebank, F.; Beek, W.; and Liem, J. 2012. Systems thinking in biology–the case for qualitative representations and reasoning1. In *Proceedings of the QR12 workshop on deep knowledge representation challenge*. Citeseer.

[Dahdul et al. 2014] Dahdul, W. M.; Cui, H.; Mabee, P. M.; Mungall, C. J.; Osumi-Sutherland, D.; Walls, R. L.; and Haendel, M. A. 2014. Nose to tail, roots to shoots: spatial descriptors for phenotypic diversity

in the biological spatial ontology. *Journal of biomedical semantics* 5(1):1.

[De Bono et al. 2015] De Bono, B.; Safaei, S.; Grenon, P.; Nickerson, D. P.; Alexander, S.; Helvensteijn, M.; Kok, J. N.; Kokash, N.; Wu, A.; Yu, T.; et al. 2015. The open physiology workflow: modeling processes over physiology circuitboards of interoperable tissue units. *Frontiers in physiology* 6.

[Degtyarenko et al. 2008] Degtyarenko, K.; De Matos, P.; Ennis, M.; Hastings, J.; Zbinden, M.; McNaught, A.; Alcántara, R.; Darsow, M.; Guedj, M.; and Ashburner, M. 2008. Chebi: a database and ontology for chemical entities of biological interest. *Nucleic acids research* 36(suppl 1):D344–D350.

[Falkenhainer and Forbus 1991] Falkenhainer, B., and Forbus, K. D. 1991. Cmpositional modeling: finding the right model for the job. *Artificial intelligence* 51(1):95–143.

[Falkenhainer and Rajamoney 1988] Falkenhainer, B., and Rajamoney, S. 1988. The interdependencies of theory formation, revision, and experimentation. In *Proceedings of the 5th International Conference on Machine Learning*, 353–366.

[Forbus 1984] Forbus, K. D. 1984. Qualitative process theory. *Artificial intelligence* 24(1):85–168.

[Fung 2013] Fung, Y.-c. 2013. *Biomechanics: mechanical properties of living tissues*. Springer Science & Business Media.

[Gidaya et al. 2016] Gidaya, N. B.; Lee, B. K.; Burstyn, I.; Michael, Y.; Newschaffer, C. J.; and Mortensen, E. L. 2016. In utero exposure to -2-adrenergic receptor agonist drugs and risk for autism spectrum disorders. *Pediatrics*.

[Herrero-Zazo et al. 2015] Herrero-Zazo, M.; Segura-Bedmar, I.; Hastings, J.; and Martínez, P. 2015. Dinto: using owl ontologies and swrl rules to infer drug–drug interactions and their mechanisms. *Journal of chemical information and modeling* 55(8):1698–1707.

[Joshi-Tope et al. 2005] Joshi-Tope, G.; Gillespie, M.; Vastrik, I.; D'Eustachio, P.; Schmidt, E.; de Bono, B.; Jassal, B.; Gopinath, G.; Wu, G.; Matthews, L.; et al. 2005. Reactome: a knowledgebase of biological pathways. *Nucleic acids research* 33(suppl 1):D428–D432.

[Kenkre, Tan, and Bloom 2013] Kenkre, J.; Tan, T.; and Bloom, S. 2013. Treating the obese diabetic. *Expert review of clinical pharmacology* 6(2):171–183.

[Kibbe et al. 2015] Kibbe, W. A.; Arze, C.; Felix, V.; Mitraka, E.; Bolton, E.; Fu, G.; Mungall, C. J.; Binder, J. X.; Malone, J.; Vasant, D.; et al. 2015. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic acids research* 43(D1):D1071–D1078.

[Klenk et al. 2012] Klenk, M.; de Kleer, J.; Bobrow, D. G.; Yoon, S.; Hanley, J.; and Janssen, B. 2012. Guiding and verifying early design using qualitative simulation. In *Proceedings of the ASME 2012 IDETC and CIE*.

[Kuipers 1994] Kuipers, B. 1994. *Qualitative reasoning: modeling and simulation with incomplete knowledge*. Cambridge, MA, USA: MIT Press.

[Li and Agarwal 2009] Li, Y., and Agarwal, P. 2009. A pathway-based view of human diseases and disease relationships. *PloS one* 4(2):e4346.

[Liebman 2010] Liebman, T. 2010. The role of cortisol and abdominal obesity in the epidemic of type 2 diabetes. *Undergraduate Research Journal for the Human Sciences* 9.

[Mayfield et al. 1998] Mayfield, J. A.; Reiber, G. E.; Sanders, L. J.; Janisse, D.; and Pogach, L. M. 1998. Preventive foot care in people with diabetes. *Diabetes Care* 21(12).

[Organization, Organization, and others 1992] Organization, W. H.; Organization, W. H.; et al. 1992. Icd-10. *International statistical classification of diseases and related health problems (10th edition). Geneva, SR: World Health Organization (WHO)*.

[Platek et al. 2009] Platek, M. E.; Shields, P. G.; Marian, C.; McCann, S. E.; Bonner, M. R.; Nie, J.; Ambrosone, C. B.; Millen, A. E.; Ochs-Balcom, H. M.; Quick, S. K.; et al. 2009. Alcohol consumption and genetic variation in mthfr and mtr in relation to breast cancer risk. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 18(9):2453.

[Rickel and Porter 1997] Rickel, J., and Porter, B. 1997. Automated modeling of complex systems to answer prediction questions. *Artificial Intelligence* 93(1):201–260.

[Smith et al. 2007] Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L. J.; Eilbeck, K.; Ireland, A.; Mungall, C. J.; et al. 2007. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology* 25(11):1251–1255.

[Vezin and Verriest 2005] Vezin, P., and Verriest, J. P. 2005. Development of a set of numerical human models for safety. In *19th International Technical Conference on the Enhanced Safety of Vehicles, Washington DC*, 6–9.

[Weld and de Kleer 1989] Weld, D., and de Kleer, J. 1989. *Readings in Qualitative Reasoning about Physical Systems*. Morgan Kaufmann.