

# Introducing Digital Doppelgängers for Healthcare Policy Analysis



Jennifer Legaspi and Shamsnaz Virani Bhada

**Abstract** Healthcare policy evaluation is a time-consuming, challenging process due to the complexity of the US healthcare system which is comprised of both public and private payers; a variety of healthcare suppliers including doctors, medical device companies, and pharmacies; and patients from different insurance coverages and socioeconomic backgrounds. Systems engineering processes are intended for complex systems and are ideal for addressing healthcare policy. Specifically, model-based systems engineering (MBSE) is used to increase traceability with its model-centric approach and can be used to increase understanding of the healthcare system. In this paper, we attempt to exploit digital twin philosophy of MBSE to understand a US healthcare system as a complex system. We focus our efforts in building a digital doppelgänger which reflects most aspects of the healthcare systems digitally, but is not an exact digital twin. The doppelgänger helps navigate around the medical privacy laws of the US healthcare system and runs some analysis on healthcare policy.

**Keywords** Model-based systems engineering · Policy development · Model-based design · Policy analysis · Healthcare · Digital twin · Virtual prototype · Emergent properties · Digital doppelgänger

## 1 Introduction

Healthcare policy implementation in the USA can take years from the identification of a need for a policy change to the implementation of the policy. Legislators must work with the current system to make changes that will not detrimentally affect the actors in the system. Most recently, healthcare reform has occurred in 2010 with the Affordable Care Act (ACA) yielding new pathways of insurance coverage (Mulligan

---

J. Legaspi (✉) · S. V. Bhada  
Worcester Polytechnic Institute, Worcester, MA, USA  
e-mail: [jlegaspi@wpi.edu](mailto:jlegaspi@wpi.edu)

2017). The implementation of the ACA in 2010 was met by strong opposition (Wei and Jarlenski 2014), but during its cycle of utilization, it garnered support. After the ACA penalty for the uninsured was repealed in 2019, only one quarter chose not to get coverage because of the repeal (Collins and Gunja 2019). Further calls for healthcare reform have been called for by various political candidates in the USA, but a 2019 survey shows that 40% of adults do not know enough about public and private insurance to have an opinion (Collins and Gunja 2019) which may be an indicator of the level of complexity of the system.

The healthcare system is a complex system that can be analyzed using systems engineering methods and approaches to be presented to an audience in a way that is easier to understand. Model-based systems engineering (MBSE) can be used to make sense of the current rules and regulations due to its inherent benefit of traceability attributed to its model-centric approach rather than a document-centric approach (Krishnan et al. 2018). Digital twins in systems engineering can also be used to address the issue of healthcare reform by accurately representing the complexity of the health records of an individual.

Typically, digital twins are implemented as a representation of a specific instance that are continually updated with respect to its twin (Madni et al. 2019). In the case of healthcare, there are rules for medical record privacy and security in the USA under the Health Insurance Portability and Accountability Act (HIPAA) (Madni et al. 2019).

Security risks that exist with a digital twin of healthcare records should be considered when using a digital twin for policy evaluation. This research addresses these security risks in healthcare reform analysis by using a digital doppelgänger instead of a digital twin.

In this case we provide a distinction between digital twin and digital doppelgänger. A digital twin of a health record would be a complete and digital version of a specific person's health records. The Meriam-Webster defines a doppelgänger as "something or someone that strongly resembles another." From this definition, we describe a digital doppelgänger in terms of health records as a synthetic version of health records that could bear strong resemblance to the health records of a person that exists in the world, but the digital doppelgänger is not related to the person's health records, nor does a change in that person's health have an effect on the synthetic health records of the digital doppelgänger. Figure 1 illustrates the differences between a digital twin and digital doppelgänger with a snapshot over multiple years of life. The living person has matched records held in the digital realm represented by a digital twin. A digital doppelgänger has similar patterns of doctor visits to the living person and may have diagnoses and treatments at age ranges considered similar with respect to health statistics. The digital doppelgänger can be viewed as a hybrid of a virtual prototype and a digital twin.

Using MBSE and digital doppelgängers or digital twins, healthcare reform can be evaluated with respect to the system requirements modeled in SysML. Digital doppelgängers as well as digital twins can enable the simulation of emergent properties in a complex system such as healthcare in the USA.

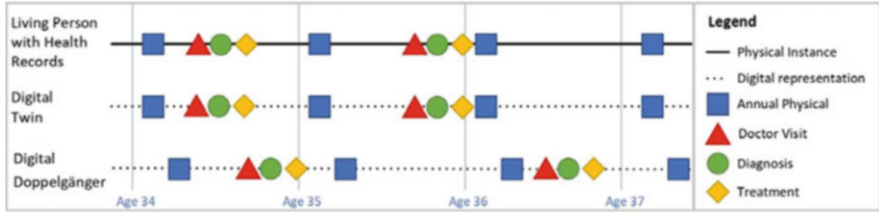


Fig. 1 Sample timeline of living person represented by a digital twin and a digital doppelgänger

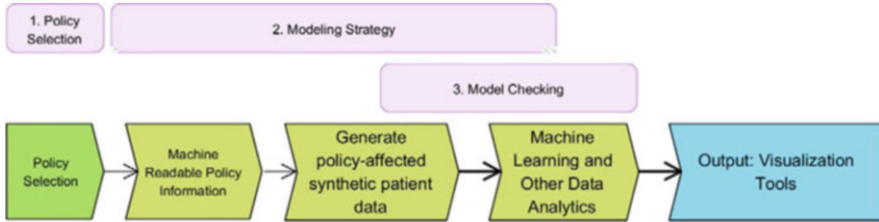


Fig. 2 Proposed methodology

## 2 Methodology

This section describes the proposed methodology behind the application of a digital doppelgänger with MBSE for healthcare policy evaluation. The methodology in Virani et al. is built upon to generate a methodology applicable to healthcare reform. The proposed change to the methodology follows the initial three steps of policy selection, modeling strategy, and model checking (U.S. Department of Health and Human Services 2013) and adds model distribution as a key step in the methodology as shown in Fig. 2. In previous research, models are intended to be used by researchers familiar with the tools (Krishnan et al. 2018). In this research, the intention is to provide a tool that can be used by patients, lawmakers, suppliers, and payers to make educated decisions regarding healthcare reform.

### 2.1 Policy Selection

The first step in the methodology is policy selection. Full applications of the methodology proposed by Virani et al. will be more rigorous for the policy selection step, but because the end goal of the methodology is to create a tool for nonexperts in the healthcare domain, policies that can potentially affect a wide range of actors in the system are prioritized over policies that may only affect a small group for the time being. Taking this into consideration, the research team selected basic, yet commonly misunderstood, aspects of health insurance such as copays and premiums

for a basic proof of concept. A full application of the methodology would require a specific and complete policy that has been implemented or proposed.

Copays or copayments are fixed amounts paid for by the patient for a covered healthcare service. Premiums are paid on monthly basis by the health insurance subscriber (Virani and Rust 2016). Patients would ideally like to have no copays or premiums while still having access to quality medical services. Suppliers, which include doctors and pharmacies, benefit from copayments because it reduces the amount of upfront costs they must pay to provide a medical service to a patient. If a supplier is not reimbursed by the payer and the patient does not pay, they will only get the amount of the copayment for the medical service. Payers benefit from premiums which go into their funds from each of their policy subscribers. If payers have no funds, they cannot reimburse suppliers for medical services supplied to patients.

The policy changes selected for focus are change in copays and change in monthly premiums. Because the example policy selected is a simple policy change rather than a complex and specific proposed implementation, the use of natural language processing is not required (U.S. Department of Health and Human Services 2013). When applying this research to a specific proposition, natural language processing will likely be required to automate the policy modeling process.

## 2.2 *Modeling Strategy*

The modeling strategy makes use of both SysML and MATLAB with digital doppelgängers. MBSE methods emphasize the importance of traceability between SysML and MATLAB. Synthea, an open-source tool created by MITRE Corporation, provides access to digital doppelgängers which are utilized to help uncover emergent properties of the healthcare system. The digital doppelgängers generated by Synthea are based on real patient datasets and medical data to generate synthetic patient data. This synthetic electronic health record (EHR) is not connected to a human system, so it does not have the privacy and security concerns that real electronic health records have (MITRE 2019).

Using SysML, we model the effects of changes in copays and premiums on the relationships in the healthcare system, further document Synthea in model form, and document any additional work completed in MATLAB. The sequence diagram represents the effects of decreased copays on healthcare system through a patient visit to the doctor. Copays and premiums have direct effects on the actors in the relationships, but they can also indirectly affect the relationships and interactions between actors in the system. Lower copays, for example, may encourage more frequent doctor visits causing longer wait times and shorter amounts of time spent with the doctor. The lower copays may also increase the likelihood of the patient being able to afford any prescriptions provided by the doctor. According to the American Medical Association (AMA), medical nonadherence – where patients do not take prescriptions as prescribed by their doctor – is common. The AMA states

that 25% of new prescriptions are not filled and 50% are not used as directed (U.S. Centers for Medicare and Medicaid Services 2019). These interactions in groups of relationships happen hundreds of millions of times annually which has a significant effect on the healthcare system overall.

Due to the broadness and complexity of the healthcare system in the USA, the SysML modeling strategy must be kept in focus by the scope of the system. The healthcare system is comprised of a range of payers, providers, patients, payer policies, as well as policies in the form of government rules and regulations. Rules and regulations can vary from state to state with some states requiring more benefits than others. By limiting the scope, the models can be better defined for the payer policies that most resemble the effects of changes in copays and premiums. In addition to limiting the scope to changes in copays and premiums, we also limit the scope by focusing on a population that resembles the state of Massachusetts.

The sequence diagram in SysML is also color coded to show how changes may either positively affect or negatively affect relationships in the system. Shown in Fig. 3, positive effects are indicated by green comment boxes. Negative effects are indicated by red comment boxes.

In addition to SysML, MATLAB is used to mathematically represent the relationships in the healthcare system. Many of the relationships are already modeled in Synthea, but aspects of the relationships that have not yet been implemented in Synthea are modeled in MATLAB. Anything modeled in MATLAB is also modeled in SysML to document functions and ensure traceability. Figures 4 and 5 show the SysML representation of a MATLAB implementation of the payers, respectively. Changes in copays and premiums were initially modeled in MATLAB until they were added to Synthea. The results in MATLAB can also be used to visualize the effects of copay and premium changes on the system. The relationships represented in MATLAB and Synthea are based on real-world data. Data from the US government as well as from Commonwealth Fund was utilized in the MATLAB models. Each function in MATLAB was modeled in SysML in order to document the purpose of the function and to provide traceability. Payers were modeled in SysML in connection to the MATLAB implementation.

At the time, payer copay data had not been implemented in Synthea, so the copays were implemented in MATLAB. Synthea models were created by MITRE Corporation and are being continually improved. While the team did not make changes to the models that were created in Synthea, changes could be requested if needed. Copay and premium data are examples of changes that were made by MITRE in parallel to this research. Recognizing the limitations of the Synthea models was important to ensuring the accuracy of the models when integrated with Synthea.

The data produced by Synthea is both a benefit and a limitation of the Synthea models by functioning a hybrid of a digital twin and a virtual prototype. We refer to this hybrid as a digital doppelgänger which we leverage to work with US medical privacy laws. In the USA, medical privacy laws such as HIPAA impose restrictions on access to medical records (Madni et al. 2019). For the purpose of analyzing health policy, restrictions on access to medical records can be prohibitive to the

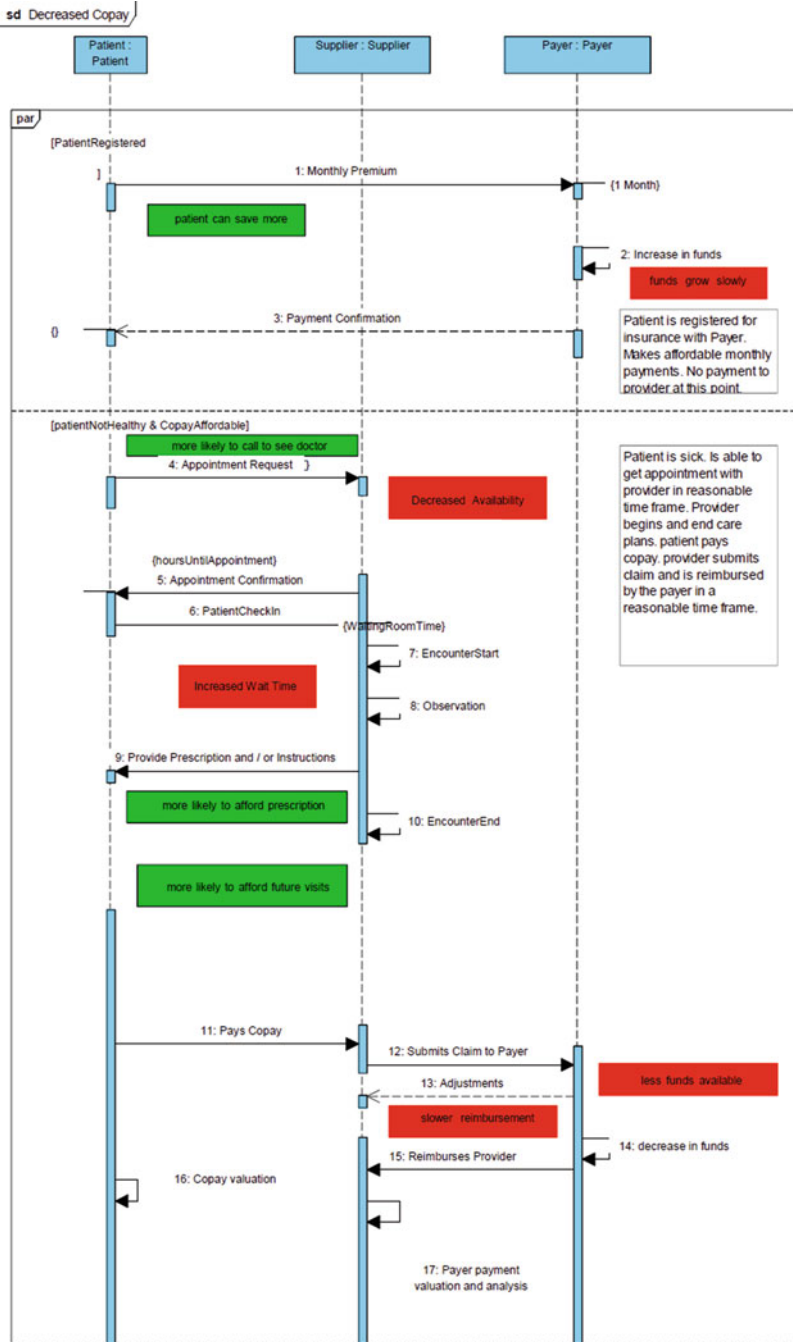


Fig. 3 Sequence diagram of patient, supplier, payer relationship

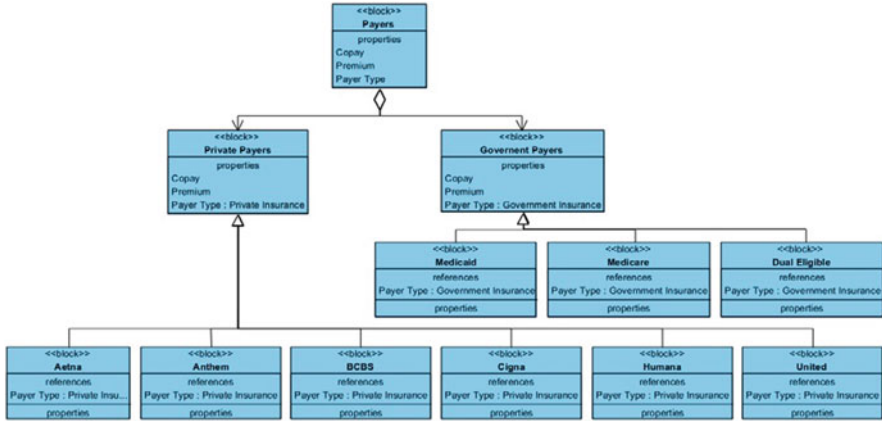


Fig. 4 Block diagram of payers

```
function copayPatient = copay(payerID)

[s, z] = size(payerID);
copayPatient = zeros(s,z);
for n = 1:s
    if payerID(n)=='b3221cfc-24fb-339e-823d-bc4136cbc4ed'; %Dual Eligible
        copayPatient(n) = 100;
    elseif payerID(n)=='7caa7254-5050-3b5e-9eae-bd5ea30e809c'; %Medicare
        copayPatient(n) = 40;
    elseif payerID(n)=='7c4411ce-02f1-39b5-b9ec-dfbea9ad3c1a'; %Medicaid
        copayPatient(n) = 60;
    elseif payerID(n)=='d47b3510-2895-3b70-9897-342d681c769d'; %Humana
        copayPatient(n) = 80;
    elseif payerID(n)=='6e2f1a2d-27bd-3701-8d08-dae202c58632'; %BCBS
        copayPatient(n) = 75;
    elseif payerID(n)=='5059a55e-5d6e-34d1-b6cb-d83d16e57bcf'; %United
        copayPatient(n) = 70;
    elseif payerID(n)=='0e582013-fcdf-3f4b-9801-c8835acf347c'; %Aetna
        copayPatient(n) = 65;
    elseif payerID(n)=='047f6ec3-6215-35eb-9608-f9dda363a44c'; %Cigna
        copayPatient(n) = 60;
    elseif payerID(n)=='42c4fca7-f8a9-3cd1-982a-dd9751bf3e2a'; %Anthem
        copayPatient(n) = 55;
    elseif payerID(n)=='b1c428d6-4f07-31e0-90f0-68ffa6ff8c76'; %NO_INSURANCE
        copayPatient(n) = 100;
    else
        copayPatient(n) = 20;
    end
end
end
```

Fig. 5 MATLAB implementation of payers

analysis. Synthea takes the benefits of a digital twin such as a high level of detail and instantiations as well as the benefits of a virtual prototype but does not represent a real person to provide digital doppelgängers. Instantiations could have similar

medical records to a real person but are not actually connected to a real person. Character strings are appended to automatically generated names to show that the instantiations are generated by Synthea and are not real people to prevent misuse of data and to prevent confusion (MITRE 2019).

### 3 Model Checking

The models have three major parts which need to be checked: SysML, MATLAB, and Synthea. The strategies for SysML and MATLAB were nearly identical. With Synthea, the model checking strategy focused on Synthea integration with MATLAB and SysML models.

In a preliminary check, the SysML diagrams and the MATLAB functions were reviewed by subject matter experts at MITRE Corporation during the model checking phase. During a complete implementation of the methodology with a specific policy instance, functions may need to be checked line by line. Modeling was completed in an iterative approach, so functionality was added incrementally and checked by experts throughout the process. With SysML, the models should be confirmed by experts who compare it to current policies. MATLAB models are also compared to policies, but these models have the added benefit of results that could be displayed. Variance from expectation could indicate the need to adjust MATLAB models. Once models are thoroughly vetted, they can be used to indicate emergent properties of the system. Accurate and sufficiently fitted models checked by experts could demonstrate unexpected effects of the complex healthcare system.

Synthea is checked by subject matter experts at MITRE Corporation. With respect to this research, SysML and MATLAB models needed to be checked for proper integration with Synthea to ensure that functionality would not be doubly implemented and that any assumptions about features in Synthea were confirmed with the MITRE team.

### 4 Conclusion

This paper discussed the use of digital twins and digital doppelgängers with MBSE constructs and methodologies to model and analyze healthcare policy. The approach builds on existing approaches but modifies them for the healthcare system. Digital doppelgängers are used to help simulate emergent properties of the healthcare system, but as electronic health records become increasingly prevalent and the security of those health records is addressed, digital twins can be used to even more accurately model the emergent properties of the healthcare system.



**Acknowledgments** We thank our colleagues at MITRE Corporation for their expertise in healthcare and Synthea, especially to Jason Walonoski and Rob Lieberthal. Additional thanks to Robi Scalfani for generating datasets from unreleased branches of Synthea.

## References

- Collins, S.R., and M.Z. Gunja. 2019. What Do Americans Think About Their Health Coverage Ahead of the 2020 Election? *Commonwealth Fund*, September 26 2019. [Online]. Available: <https://www.commonwealthfund.org/publications/issue-briefs/2019/sep/what-do-americans-think-health-coverage-2020-election>. Accessed 17 Oct 2019.
- Krishnan, R., S. Virani, and R. Gasoto. 2018. Discovering toxic policies using MSBE constructs. In *Disciplinary Convergence in Systems Engineering*, Redondo Beach.
- Madni, A.M., C.C. Madni, and S.D. Lucero. 2019. Leveraging Digital Twin Technology in Model-Based Systems Engineering. *Systems 7*: 7.
- MITRE. *Synthea Empowers Data Driven Health IT*, MITRE, 2019. [Online]. Available: <https://synthetichealth.github.io/synthea/#about-landing>. Accessed Sept 2019.
- Mulligan, J. 2017. The Problem of Choice: From the Voluntary Way to Affordable Care Act Health Insurance Exchanges. *Social Science and Medicine* 181: 34–42.
- U.S. Centers for Medicare & Medicaid Services. *Glossary* [Online]. Available: <https://www.commonwealthfund.org/publications/issue-briefs/2019/sep/what-do-americans-think-health-coverage-2020-election>. Accessed 17 Oct 2019.
- U.S. Department of Health and Human Services. 2013. Summary of the HIPAA Security Rule, U.S. Department of Health and Human Services, 26 July 2013. [Online]. Available: <https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html>. Accessed 17 Oct 2019.
- Virani, S., and T. Rust. 2016. Using Model Based Systems Engineering in Policy Development: A Thought Paper. In *Conference on Systems Engineering Research*.
- Wei, Z., and M. Jarlenski. 2014. The Politics of Opposition to the Enactment of the Patient Protection and Affordable Care Act in the United States. *International Critical Thought* 4 (2): 208–220.