

A Mathematical Framework to Estimate Chronic Hepatitis B Prevalence and Undiagnosed Proportion in Canada, 2007-2021

Julien Smith-Roberge, Farinaz Forouzannia, Abdullah Hamadeh, Zeny Feng,
Nashira Popovic, William W. L. Wong

Submitted to: JMIR Public Health and Surveillance
on: September 16, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	4
Supplementary Files.....	21
Figures	22
Figure 1.....	23
Figure 2.....	24
Figure 3.....	25
Figure 4.....	26
Figure 5.....	27
Figure 6.....	28
Multimedia Appendixes	29
Multimedia Appendix 1.....	30
Multimedia Appendix 2.....	30
Multimedia Appendix 3.....	30

A Mathematical Framework to Estimate Chronic Hepatitis B Prevalence and Undiagnosed Proportion in Canada, 2007-2021

Julien Smith-Roberge¹ PhD; Farinaz Forouzannia¹ PhD; Abdullah Hamadeh¹ PhD; Zeny Feng² PhD; Nashira Popovic³ PhD; William W. L. Wong¹ PhD

¹ School of Pharmacy University of Waterloo Kitchener CA

² Department of Mathematics and Statistics University of Guelph Guelph CA

³ Centre for Communicable Diseases and Infection Control Public Health Agency of Canada Ottawa CA

Corresponding Author:

William W. L. Wong PhD

School of Pharmacy
University of Waterloo
10A Victoria St. S.
Kitchener
CA

Abstract

Background: Chronic hepatitis B (CHB) remains a significant health burden for at least a hundred thousand Canadians. The government of Canada has endorsed the Global Strategy to eliminate hepatitis as a public health threat by 2030, but effectively targeting public health interventions is complicated by the silent nature of the disease, which can remain asymptomatic for decades.

Objective: This study develops a framework to estimate the prevalence of chronic hepatitis B and the proportion of the infected population that remains undiagnosed. We apply the proposed method to national data from Canada, from 2007 to 2021.

Methods: We infer the prevalence and undiagnosed proportion of CHB by fitting a mathematical state-transition model, based on the natural history of chronic hepatitis B, to observed CHB-related events.

Results: We estimate the national prevalence of CHB in Canada in 2021 to be 0.308% (95% CI: 0.252% - 0.363%). The corresponding percentage of undiagnosed cases was estimated to be 33.9% (95% CI: 22.7% - 45.1%).

Conclusions: The estimates of CHB prevalence obtained via our method are in line with previous estimates obtained from national seroprevalence studies. More specialized estimates, stratified by province or age cohort, may be achievable with detailed health administrative data.

(JMIR Preprints 16/09/2024:66309)

DOI: <https://doi.org/10.2196/preprints.66309>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/66309>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/66309>

Original Manuscript

Original Paper

A Mathematical Framework to Estimate Chronic Hepatitis B Prevalence and Undiagnosed Proportion in Canada, 2007-2021

Abstract

Background: Chronic hepatitis B (CHB) remains a significant health burden for at least a hundred thousand Canadians. The government of Canada has endorsed the Global Strategy to eliminate hepatitis as a public health threat by 2030, but effectively targeting public health interventions is complicated by the silent nature of the disease, which can remain asymptomatic for decades.

Objective: This study develops a framework to estimate the prevalence of chronic hepatitis B and the proportion of the infected population that remains undiagnosed. We apply the proposed method to national data from Canada, from 2007 to 2021.

Methods: We infer the prevalence and undiagnosed proportion of CHB by fitting a mathematical state-transition model, based on the natural history of chronic hepatitis B, to observed CHB-related events. Data for the calibration were obtained from the Public Health Agency of Canada and Statistics Canada.

Results: We estimate the national prevalence of CHB in Canada in 2021 to be 0.483% (95% CI: 0.443% - 0.535%). The corresponding percentage of undiagnosed cases was estimated to be 54.3% (95% CI: 50.9% - 58.9%).

Conclusions: The estimates of CHB prevalence obtained via our method are in line with previous estimates obtained from national seroprevalence studies. More specialized estimates, stratified by province or age cohort, may be achievable with detailed health administrative data.

Keywords: hepatitis B; prevalence; undiagnosed proportion; mathematical modeling; estimate; Canada

Introduction

Chronic hepatitis B (CHB) is a silent disease. In its early stages, which can last decades, it is largely asymptomatic. Because of this, those with the disease can remain undiagnosed for years until they are alerted to their condition by late-stage complications: decompensated cirrhosis, hepatocellular carcinoma (HCC), liver failure. It was estimated that hepatitis B caused 1.1 million deaths worldwide in 2022 [CITATION Wor24 \l 4105].

In 2016, Canada endorsed the Global Strategy to eliminate hepatitis as a public health threat by 2030 [CITATION Gov16 \l 4105]. It has already made progress towards this goal; a seroprevalence study published in 2013 estimated the national prevalence of CHB to be 0.4%, well below the global average [CITATION Rot13 \l 4105]. This is in large part thanks to Canada's robust vaccination program. In 2021, vaccine coverage was 89% for those over 14 years of age [CITATION Gov23 \l 4105]. However, many challenges remain: hepatitis B was one of the most frequently refused vaccines in 2021, third behind only varicella and rotavirus, and the lack of a nationally standardized vaccine schedule means that children who move across provincial lines may miss vaccination due to conflicting schedules [CITATION Gov23 \l 4105][CITATION Gov231 \l 4105]. Moreover, this vaccination program leaves several key populations under-protected, notably immigrants from countries with high hepatitis B prevalence and children exposed prior to vaccination, a risk that is elevated in provinces that do not have an infant vaccination program. The same 2013 seroprevalence study found that the CHB prevalence among the foreign-born population was 1.6%, four times the

national average [CITATION Rot13 \l 4105], and 90% of unvaccinated infants born to HBV-infected mothers will develop CHB within 6 months [CITATION Tra09 \l 4105]. The risk of contracting CHB also increases with exposure to infected blood and other body fluids, such as through needle sharing or unprotected sex [CITATION Wor24 \l 4105].

These at-risk groups comprise a significant portion of Canada's population. According to the most recent Canadian census published in 2021 [CITATION Sta23 \l 4105], immigrants represent 23% of the country's total population. Many of these originate from regions where hepatitis B virus (HBV) is either highly or moderately endemic, notably China and the Philippines, which, along with India, have been the top three sources of immigrants to Canada since 1991. As a result, a significant proportion of the CHB population in Canada is comprised of immigrants. Furthermore, Canada's vaccination program leaves a significant portion of the population under-protected. Many of the province is still perform adolescent vaccination instead of birth dose vaccination [CITATION Ont23 \l 4105]. Other gaps exist in the country's prevention strategy. For example, in Ontario, prenatal screening for hepatitis B virus (HBV) was not universal, with a coverage of 92.7% between 2012 and 2016. These gaps resulted in at least 139 children born between 2003 and 2013 being infected with HBV before being eligible for adolescent vaccination [CITATION Bio20 \l 4105].

Meeting Canada's elimination goals will require further evidence-based interventions, such as screening and community outreach, rooted in robust estimates of CHB prevalence to track the impact and the efficacy of these investments in public health. In this paper, we present a model-based framework for estimating the prevalence and undiagnosed proportion of CHB. Our approach follows in the line of similar back-calculation methods used for other diseases with long incubation periods, such as hepatitis C [CITATION Ham19 \l 4105] [CITATION Tru14 \l 4105] and HIV [CITATION Lon92 \l 4105]. Similar methods have also been used to estimate the prevalence and undiagnosed fraction of COVID-19 [CITATION Ham21 \l 4105]. To our knowledge, the present study represents the first attempt to apply such back-calculation techniques to chronic hepatitis B.

Methods

In this section, we describe the mathematical framework we use to estimate the prevalence of CHB in populations of interest. We developed a state-transition model of CHB progression based on the natural history of the disease, building on earlier work in [CITATION Kim22 \l 4105]. The model includes several late-stage clinical health states, such as decompensated cirrhosis and HCC, which are more readily diagnosed than the asymptomatic stages of CHB. Indeed, we assume that the symptoms of these late-stage clinical health states are severe enough that they will be diagnosed within one year. Using a Bayesian calibration method based on the Metropolis-Hastings algorithm, we used publicly available diagnosis data for these late-stage complications, in combination with diagnosis data for CHB itself, to back-calculate the prevalence and undiagnosed fraction of CHB.

Natural History Model

We developed a state transition model to describe the infection, progression, and treatment process, following the natural history of CHB. A schematic of this model is presented in (Figure 1). We assume that people infected with CHB will remain in exactly one state for the duration of a given calendar year. Once per year, individuals may transition to a different state based on probabilities derived from the parameters in (Table 1) and (Appendix 1). Some of these state transitions, such as the transition from an undiagnosed state to a diagnosed state, correspond to observable events recorded by public health officials. Our goal is to infer the number of individuals in each state by calibrating the model to reproduce the time evolution of these observables.

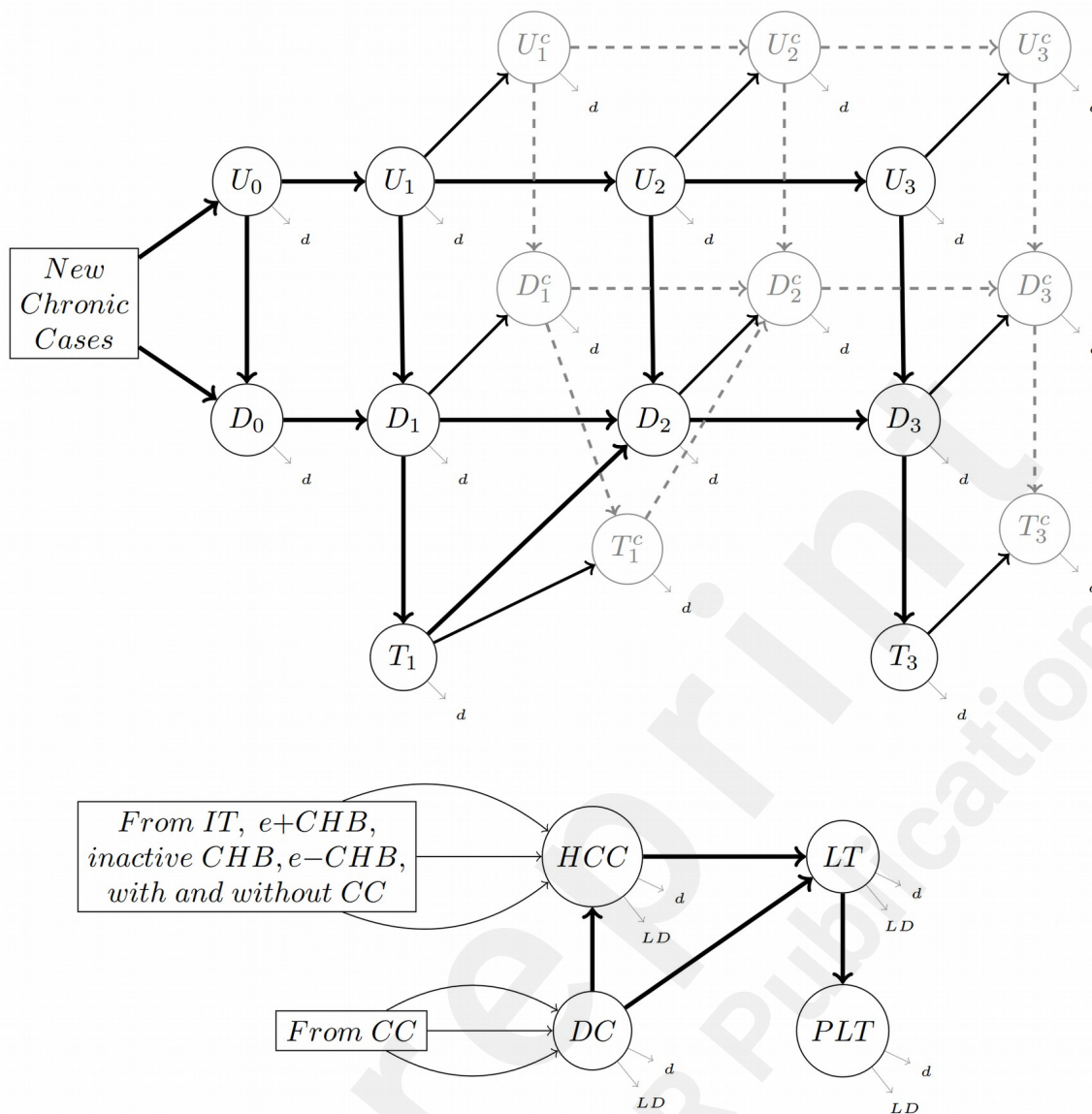


Figure 1: A schematic representation of the state-transition model.

The structure of our state-transition model is depicted in (Figure 1). For clarity, the early stages in the top portion of the diagram are organized along x-, y-, and z-axes based on their nature. The states are organized in rows and denoted by U , D , and T , corresponding to the cascade of care, specifically, the undiagnosed/untreated states, for those who have CHB but are unaware of their infection; the diagnosed/untreated states, for those with a positive CHB diagnosis who have not initiated treatment; and diagnosed/treated states, for those receiving treatment for their CHB infection. Each of these basic states are further subdivided into sub-stages, organized into columns, that represent the natural history of CHB. These sub-stages are denoted using a subscript, i :

- $i = 1$: Chronic HBeAg+ Hepatitis B (hepatitis B surface antigen (HBsAg): positive, hepatitis B e-antigen (HBeAg): positive, viral load: high, ALT level: high)
- $i = 2$: Inactive Hepatitis B (HBsAg: positive, HBeAg: negative, viral load: low, ALT level: normal)
- $i = 3$: Chronic HBeAg- Hepatitis B (HBsAg: positive, HBeAg: negative, viral load: high, ALT level: high)

Finally, each of these states can exist with or without compensated cirrhosis, which when present is represented with a superscript C . In order to cover the whole natural history of CHB, we also include

two states, U_0 and D_0 , corresponding to the immune tolerant state (HBsAg: positive, HBeAg: positive, viral load: high, ALT level: normal).

People infected with CHB may progress into advanced stages of liver disease, which are depicted in the bottom half of (Figure 1). These advanced stages are categorized into the following states:

- HCC : hepatocellular carcinoma
- DC : decompensated cirrhosis
- LT : liver transplant
- PLT : post-liver transplant

We define the i state to last at most one year. Transplant recipients who have survived more than one year after their liver transplantation enter the PLT state. Additionally, we assume that individuals in these later states have a probability of mortality greater than the background all-cause mortality. This is represented by arrows to a liver death state, LD . Background mortality is represented by an arrow to a death state, d . A full list of model states is included in (Appendix 2).

Finally, we assume that in the year t there are $u(t)$ new CHB infections. These new infections include observed infections that are diagnosed within the first year, in which case they will be directed to the state D_0 , as well as unobserved cases that remain undiagnosed and go to U_0 .

Model Transition Probabilities

The transition probabilities of the model depend on a mix of known and unknown parameters. The known parameters are obtained from the literature and are presented in (Appendix 1). The unknown parameters are presented in (Table 1). We will estimate these unknown parameters using our calibration procedure, which is described in later sections of this paper.

Table 1: Unknown Model Parameters. The minimum and maximum constrain the range explored during the calibration process.

Parameter	Description	Min	Max
d_{lr}	Annual probability of CHB diagnosis	0.005	0.5
d_{cc}	Annual probability of CC-specific diagnosis	0	0.1
d_{aa}	Probability of newly infected individuals being diagnosed with CHB within the first year	0	0.3
μ_{lr}	Annual CHB incidence	0	0.1
r_{hcc0}	Adjustment to CHB-related HCC risk	0	1

Because CHB is initially asymptomatic, we assume that the probability of receiving a diagnosis is constant across all the undiagnosed states U_i . In principle, these diagnosis probabilities can vary significantly from year to year. As a simplifying approximation, we assume that this trend is linear over the period being modeled, with a diagnosis probability of d_l in the first year being modeled, t_{init} , and a final diagnosis probability of d_r in the last year modeled, t_{fin} . For notational convenience, we denote this time-dependent quantity $d_{lr}(t)$. For undiagnosed states with compensated cirrhosis, we add an annual probability of CC-specific diagnosis, d_{cc} , to allow for the fact that compensated cirrhosis may likely produce symptoms, which increases the probability of diagnosis. Thus, the diagnosis probability for an individual in one of the U_i^C states would be $d_{lr}(t) + d_{cc}$. Finally, we apply a different diagnosis probability for the first year of infection, d_{aa} . In any given year, $d_{aa}u(t)$ new cases will be added to the D_0 state, with the remaining $(1 - d_{aa})u(t)$ cases being added to the U_0

state. This first-year diagnosis probability allows us to account for CHB-infected individuals who seek screening following high-risk behaviour, or immigrants who disclose their health status on entering the country.

Much like the diagnosis probability, we also assume that CHB incidence, μ , is a linear function of time. Following the same notational convention as above, the number of new cases in a given year is given by $u(t) = \mu_{lr}(t) Pop(t)$, where $Pop(t)$ is the population of the cohort being studied.

The final calibration parameter, r_{hcc0} , is an adjustment parameter that aims to account for uncertainty in the probability of developing HCC. The annual probabilities of developing HCC, as reported in the literature, are compiled in (Appendix 1).

Finally, we note that these model parameters, both known and unknown, represent the underlying probabilities associated to any basic state transition. In principle, simulating the full state transition model represented by (Figure 1) would also require knowledge of joint probabilities. In order to obtain a tractable model, we make the simplifying assumption that each of these basic events are independent.

Model Assumptions

We highlight some key simplifying assumptions made during the development of the model.

- Incidence rate is a linear function of time, and does not explicitly depend on other factors, such as vaccine uptake.
- The age distribution of the CHB population matches that of the general population. In particular, age-related quantities, such as the background mortality rate, are the same in both populations.
- There is no net migration of infected individuals after arrival. Moreover, because CHB has no readily available cure (spontaneous loss of HBsAg is possible, but rare), we assume that the number of cases can only decrease by death.
- Treatment uptake and efficacy remained constant over the observed period, as did survival rates of late-stage complications.

Model Dynamics

For each of the model states (U_0 , D_0 , U_1 , ..., HCC , DC , i , and PLT), the state-transition model generates a time series for the expected number of individuals in each state. We denote these by $U_0(t)$, $D_0(t)$, ..., and $PLT(t)$, for $t = t_{init}, \dots, t_{fin}$. Additionally, we record the number of individuals transitioning between states. These simulated time series—the number of individuals in each state and the number of individuals transitioning between states—are used to calculate four *observables*, which are used to calibrate the model. The first of these is the number of people receiving treatment, $T_1(t) + T_3(t) + T_1^C(t) + T_3^C(t)$. Based on clinical guidelines [CITATION Ter18 \l 4105], treatment is mainly offered to infected individuals in HBeAg+ and HBeAg- states, and thus our model does not include an inactive treatment state, T_2 . The remaining observables are the number of CHB diagnoses, the number of HCC diagnoses, and the number of late HCC diagnoses, which we define as any transition from an undiagnosed state, U_i^{-iC} , to the HCC state. These diagnosis events correspond to transitions between undiagnosed states to diagnosed states. The details of these computations can be found in (Appendix 2).

Data Sources

The observables described above are compared to three data sources: the annual number of HCC

diagnoses, obtained from Statistics Canada [CITATION Sta24 \l 4105], which we will denote $Y_{HCCTotal}(t)$; the annual number of chronic hepatitis B diagnoses, obtained from the Public Health Agency of Canada's (PHAC) Notifiable Diseases Online [CITATION Pub24 \l 4105], denoted $Y_{CHB}(t)$; and CHB treatment data, which is based on prescription data obtained from IQVIA via collaboration with PHAC, denoted $Y_{Treat}(t)$. The CHB and treatment data can be directly compared to their corresponding observable, but the other two observables require additional assumptions on the data before they can be used.

The HCC figures from Statistics Canada include all instances of HCC, regardless of cause. Following [CITATION Kim22 \l 4105], the estimated number of CHB-related HCC cases is 9% of total HCC cases, $Y_{HCC}(t) = 0.09 \cdot Y_{HCCTotal}(t)$.

In [CITATION Lap21 \l 4105], an estimated 1.07% of CHB diagnoses (198 of 18434) occur within ± 6 months of an HCC diagnosis. Thus, we assume the number of late HCC diagnoses per year is given by

$$Y_{LateHCC}(t) = 0.0107 \cdot Y_{CHB}(t).$$

Reporting HBV cases is mandatory. Therefore, we can be confident that the Notifiable Diseases Online database accurately reflects the number of HBV diagnoses. However, although HBV diagnoses must be reported, the chronic vs acute distinction is not reportable in all provinces and territories. Unspecified cases were apportioned based on the chronic-to-acute ratio observed in the specified cases. We note that although the rate of HBV diagnosis per se may be biased (not all people who are infected will seek medical attention, and therefore the Notifiable Diseases database might systematically underrepresent populations less likely to interact with the healthcare system), the structure of our mathematical model accounts for this possibility. Bias or underreporting in the diagnoses of late-stage complications represents a greater source of uncertainty in the model. As such, the HCC data from Statistics Canada represents the most significant source of data-related uncertainty in our analysis. The data for model calibration are available in (Appendix 3).

Model Calibration

The model contains several unknown quantities that need to be inferred via model calibration. These come in two broad categories: the unknown parameters from (Table 1), and the initial population sizes for each of the model states.

If the initial population distributions in the year t_{init} deviate too much from the true values, the calibration converges to a spurious local optimum. In order to avoid this, we employ a two-stage calibration process. We begin by assuming a prevalence of 0.4%, in line with earlier seroprevalence results [CITATION Rot13 \l 4105], and distribute these among the states in the model based on percentages used in previous modeling studies [CITATION Won11 \l 4105]. We then calibrated the model against data from a subset of provinces—Alberta, BC, Ontario, and Saskatchewan—whose diagnosis, treatment, and HCC data were the most complete, and selected the population distribution that yielded the closest match to their data. These population distributions then served as our initial distribution in all subsequent calibrations.

In order to obtain an initial estimate of incidence, μ_{lr} , we assume the system is in approximate equilibrium and apply the same 0.4% seroprevalence estimate. This yields a preliminary estimate of $\mu_{lr} \approx 0.00004$.

In principle, the early diagnosis parameter, d_{aa} , can take any value between zero and one, though in practice we know that values close to 1 are unrealistic. We restrict our search range to $0 \leq d_{aa} \leq 0.3$ [CITATION Lap21 \l 4105].

Calibration Algorithm

Following [CITATION Ham19 \l 4105], we employ a two-stage calibration algorithm, which is designed to account for the uncertainty in the literature-derived parameters. Let $v^{lit} = [q_{01}, \dots, d_{ld4}]$ be a set of the literature-derived model parameters, i.e. a set of parameters sampled from the ranges in (Appendix 1). Similarly, let $v^{cal} = [d_l, \dots, r_{hcc0}, U_0(t_{init}), \dots, PLT(t_{init})]$ denote a set of the calibration parameters, i.e. a set of the unknown model parameters from (Table 1) and a set of initial population estimates. Taken together, a single pair (v^{lit}, v^{cal}) determines the initial conditions and transition probabilities, and thus fully specifies the behaviour of the state transition model. The goal of our calibration is to generate pairs that yield a good fit with the available data.

Our two-stage calibration algorithm is described as follows:

Stage 1: We begin by generating k sets of the literature-derived parameters,

$[v_1^{lit}, \dots, v_k^{lit}]$, by sampling uniformly from the ranges in (Appendix 1). For each sample v_i^{lit} , we calibrate the model using the Metropolis-Hastings MCMC algorithm [CITATION Has70 \l 4105], using the likelihood function described in (Appendix 2), until the Markov chain converges to a stationary distribution, $P(v^{cal} \vee Y, v_i^{lit})$. Discarding the burn-in, we obtain a set of n samples for the unknown parameters, $[v_1^{cal}, \dots, v_n^{cal}]$, sampled from this distribution.

Stage 2: The samples generated in the first stage are conditioned on a choice of v_i^{lit} . Thus, in order to obtain a full set of model parameters, we perform a second sampling step. We first sample uniformly from the literature-derived parameters obtained in step 1, $[v_1^{lit}, \dots, v_k^{lit}]$, and then sample uniformly from the associated distribution obtained via the MCMC algorithm. The result is a single combined set of parameters, (v^{lit}, v^{cal}) , sampled from the distribution $P(v^{lit}, v^{cal} \vee Y)$.

For each pair (v^{lit}, v^{cal}) derived in this way, the model produces a time series for each of the model states, $U_0(t)$, $D_0(t)$, ..., $PLT(t)$. In order to describe the typical behaviour of our model, we sample multiple pairs (v^{lit}, v^{cal}) and compute the means of the associated time series, $\dot{U}_0(t)$, $\dot{D}_0(t)$, ..., $\dot{PLT}(t)$, as well as the corresponding standard deviations.

We obtain a prevalence estimate by summing over all model states and dividing by the total population size.

$$CHB\ prevalence(t) = [\dot{U}_0(t) + \dot{D}_0(t) + \dots + \dot{PLT}(t)] / Pop(t)$$

The undiagnosed fraction is computed as the ratio of all undiagnosed states and the total CHB.

$$undiagnosed\ fraction(t) = \frac{\dot{U}_0(t) + \dot{U}_1(t) + \dots + \dot{U}_3^c(t)}{\dot{U}_0(t) + \dot{D}_0(t) + \dots + \dot{PLT}(t)}$$

Using this two-stage calibration algorithm, we first estimate the national prevalence. We then conduct a stratified analysis for male and female populations. Finally, we conduct a one-way

sensitivity analysis to explore the uncertainty associated to these estimates.

Results

After completing the calibration process outlined above, the algorithm identified a set of model parameters and initial conditions that can closely match with the observed statistics on CHB diagnosis and HCC diagnosis (Figure 2). In order to produce such a fit, the national prevalence of chronic hepatitis B in 2021 was estimated to be 0.483% (95% CI: 0.443% - 0.535%), while the percentage of chronic hepatitis B cases undiagnosed in 2021 was estimated to be 54.3% (95% CI: 50.9% - 58.9%) (Figure 2). The model had a deviance information criterion of 53.9.

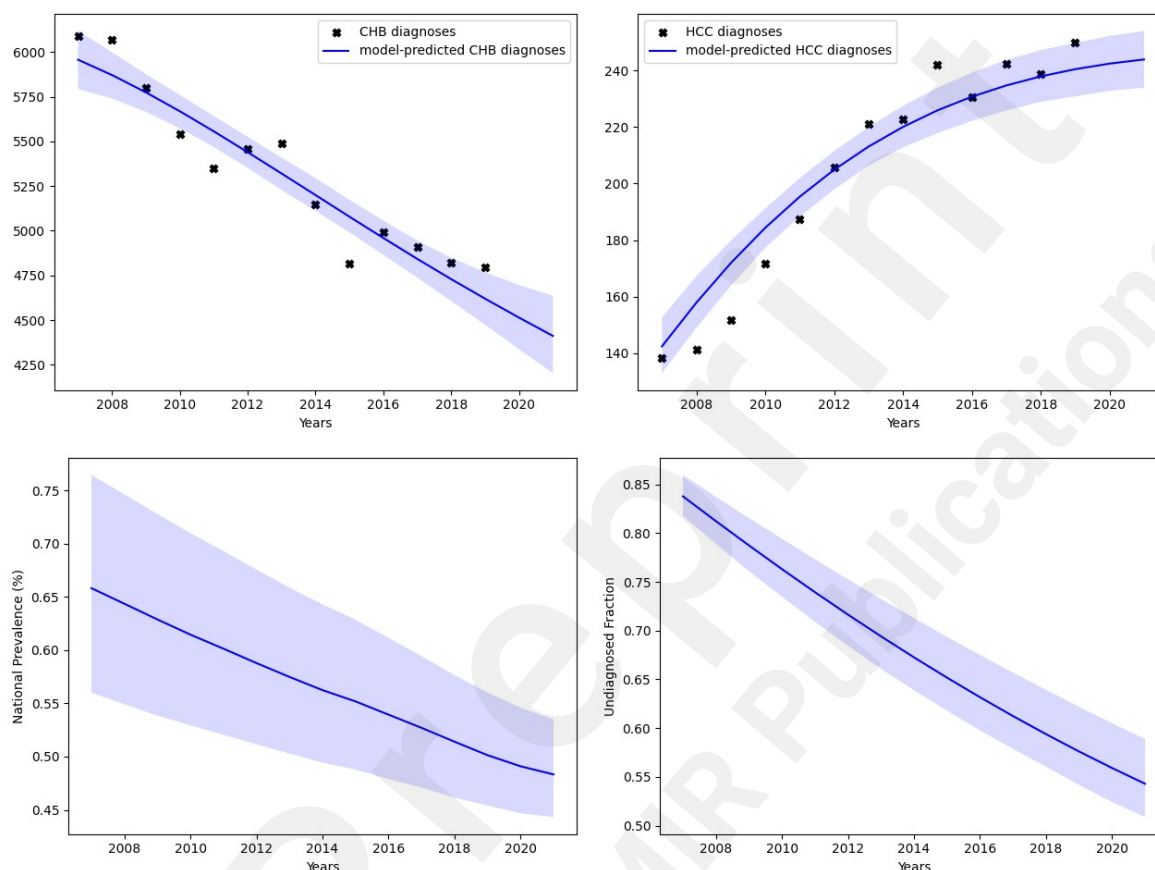


Figure 2: Estimates of the national CHB prevalence across all cohorts. The top two plots show the model fit compared to data for CHB diagnoses (top left) and CHB-attributed HCC diagnoses (top right). The bottom two plots accordingly display the estimates of prevalence (bottom left) and of the undiagnosed proportion (bottom right).

Stratified Analyses

We conducted a second calibration stratified by sex. Again, the algorithm identified a set of model parameters and initial conditions that closely match the observed statistics on CHB diagnosis and HCC diagnosis for the male population (Figure 3) and the female population (Figure 4). In order to produce these fits, the stratified analyses suggest a slightly higher prevalence of 0.663% (95% CI: 0.572% - 0.816%) in the male population, while the percentage of chronic hepatitis B cases undiagnosed in 2021 was estimated to be 60.0% (95% CI: 54.8% - 67.0%) (Figure 3). In the female population, we obtain a slightly lower prevalence of 0.382% (95% CI: 0.347% - 0.424%), while the percentage of chronic hepatitis B cases undiagnosed in 2021 was estimated to be 48.8% (95% CI: 44.2% - 54.0%) (Figure 4). The male and female fits had deviance information criteria of 9.0 and 73.2 respectively. The differences in prevalence obtained from this calibration process are consistent with the known sexual dimorphism in CHB [CITATION Bro22 \l 4105].

In addition to stratifying by sex, we also attempted an analysis based on age-cohort. However,

because late-stage complications are highly skewed towards older individuals, the dataset for the younger cohort was too small and noisy to make reliable inferences. We intend to address this in future work with better health administrative data.

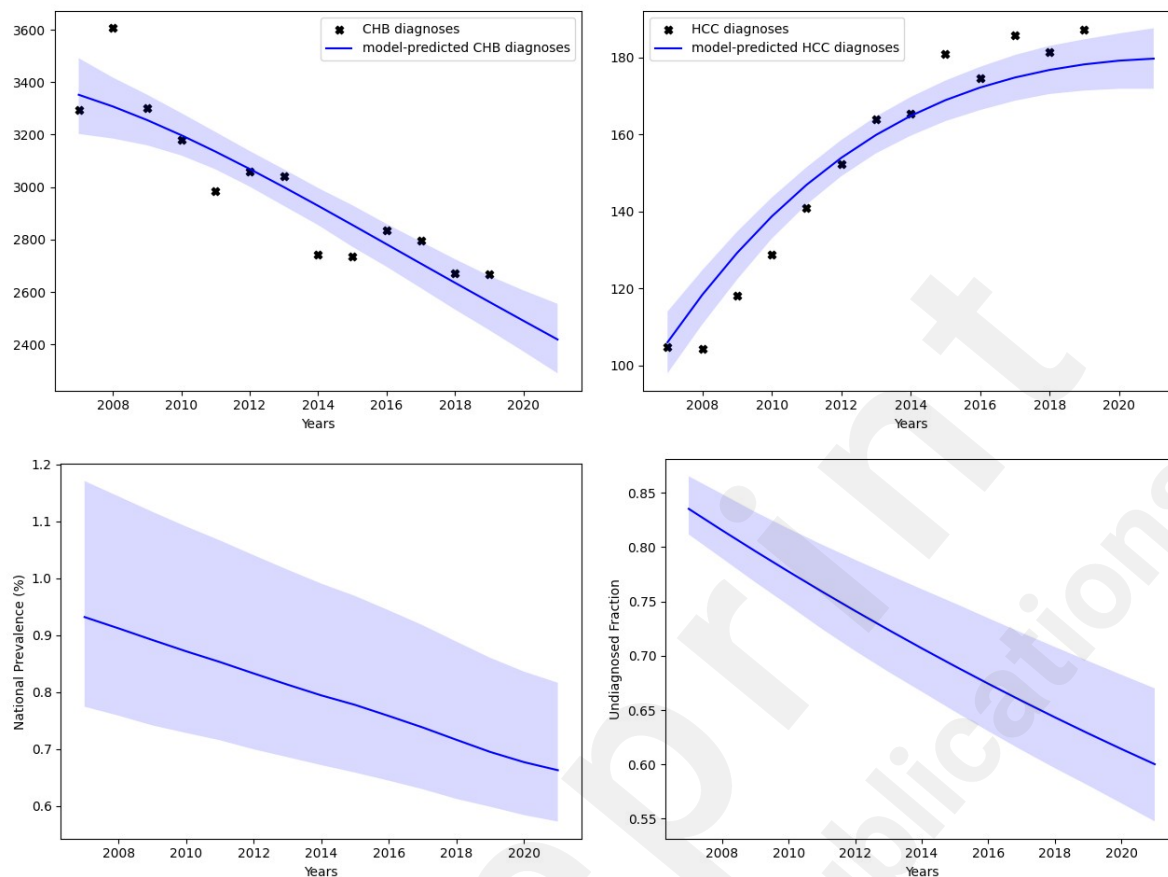


Figure 3: Estimates of the CHB prevalence in males. The top two plots show the model fit compared to data for CHB diagnoses (top left) and CHB-attributed HCC diagnoses (top right). The bottom two plots accordingly display the estimates of prevalence (bottom left) and of the undiagnosed proportion (bottom right).

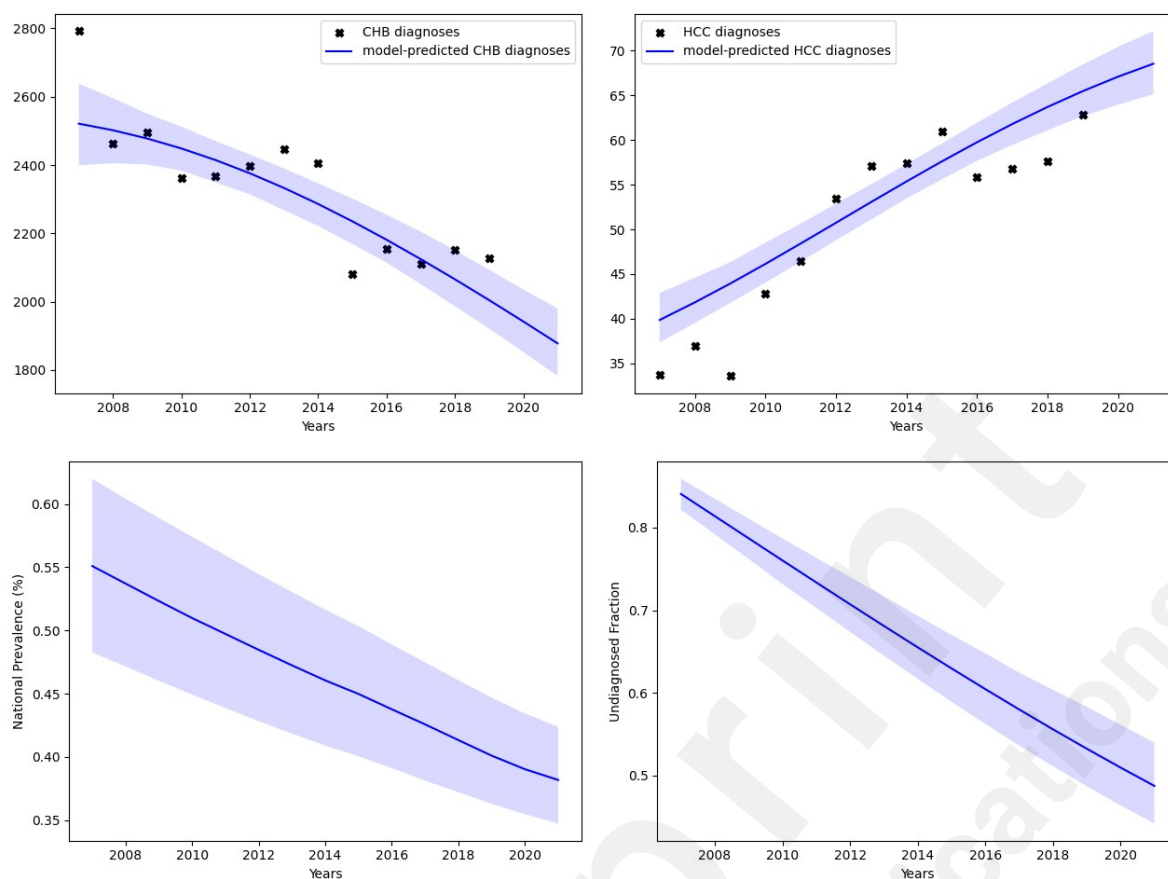


Figure 4: Estimates of the CHB prevalence in females. The top two plots show the model fit compared to data for CHB diagnoses (top left) and CHB-attributed HCC diagnoses (top right). The bottom two plots accordingly display the estimates of prevalence (bottom left) and of the undiagnosed proportion (bottom right).

Sensitivity Analysis

We performed a one-way sensitivity analysis of the calibrated model, varying the parameters in (Table 1) to determine which components of the model caused the greatest variation in the estimates of the prevalence and undiagnosed fraction. Parameters were varied by $\pm 50\%$, and we measured the resulting change in the prevalence and diagnosed fraction, averaged over the time period studied (2007-2021). The results are presented in (Figure 5) and (Figure 6).

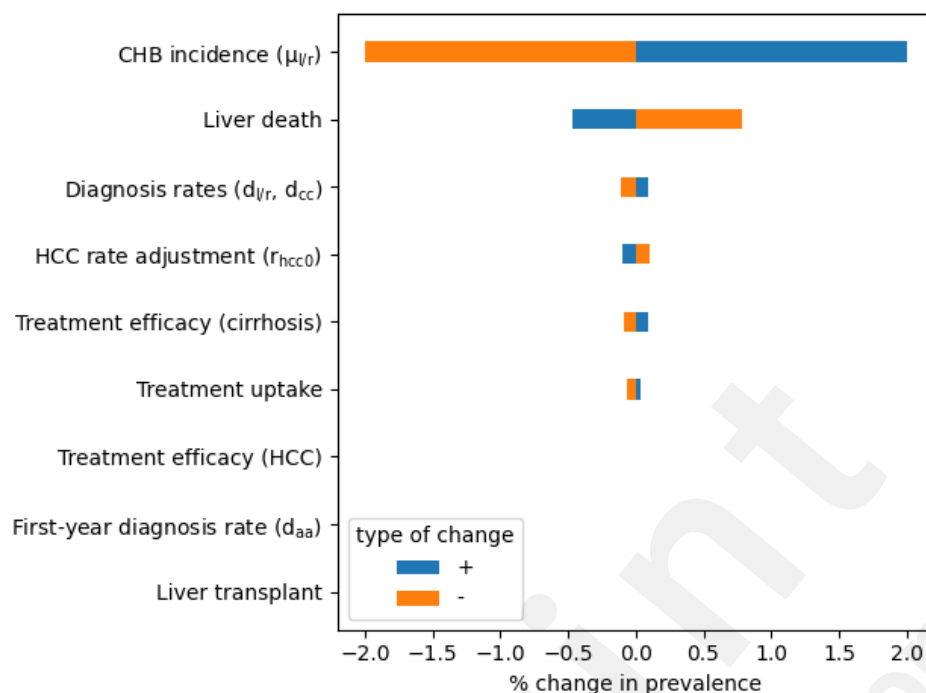


Figure 5: One-way sensitivity analysis for the estimated prevalence of CHB.

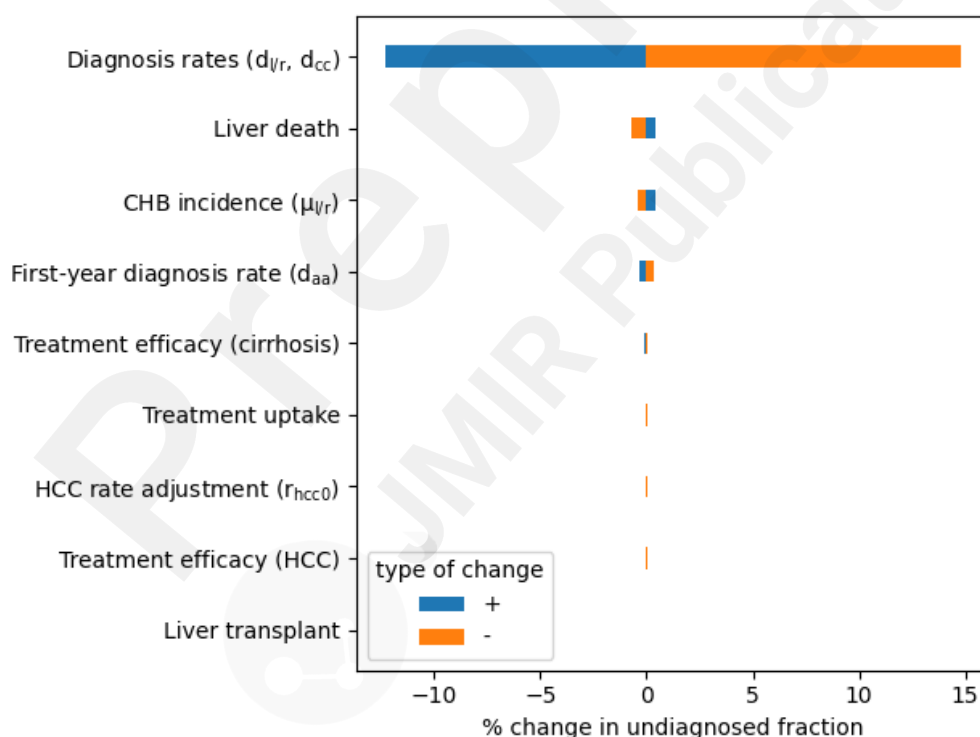


Figure 6: One-way sensitivity analysis for the estimated undiagnosed proportion of CHB infections.

As expected, the yearly CHB incidence has the largest impact of prevalence. Other parameter changes impact prevalence indirectly by changing the mortality rate. This effect was seen most strongly in the probability of liver death associated with late-stage complications. Changing the diagnosis probabilities or HCC incidence also increased prevalence, presumably because of their downstream effects on CHB-related mortality. Similarly, increasing the treatment uptake and treatment effectiveness increased prevalence due to the associated drop in mortality. The effects d_{aa}

on prevalence was negligible.

Changes in the diagnosis probabilities had the largest effect on the undiagnosed fraction, followed by the probability of liver death, CHB incidence, and d_{aa} , which had comparable effects. Changes in events downstream of diagnosis, such as HCC incidence or treatment effectiveness, did not change the overall prevalence enough to have a significant impact on the undiagnosed proportion.

Discussion

Principal Results

CHB has a long incubation period, often progressing undetected until the onset of late-stage complications. This creates significant uncertainties in the prevalence and undiagnosed proportion and poses challenges to policymakers who wish to measure the efficacy of public health interventions or track progress towards elimination targets. We have proposed a model-based method to estimate these quantities. Using limited CHB and HCC diagnosis data, we estimate prevalence in 2021 to be 0.483% (95% CI: 0.443% - 0.535%). These results are in line with previous 0.4% seroprevalence estimates [CITATION Rot13 \l 4105]. Additionally, the proposed method provides estimates of the fraction of CHB cases that remain undiagnosed, in this case 54.3% (95% CI: 50.9% - 58.9%). Again, this is in line with previous estimates. The same Canadian seroprevalence study indicated that 46% of respondents who tested positive for a present HBV infection knew of their infected status.

These numbers are comparable to estimates obtained in other countries. In 2022, approximately 205,549 individuals in Australia were living with CHB, representing 0.78% of the population. It is estimated that 27.9% of these remain undiagnosed [CITATION Aus22 \l 4105]. The prevalence of hepatitis B varies across EU/EEA countries, with estimates ranging from 0.1% in Ireland to 4.4% in Romania [CITATION Eur16 \l 4105]. The proportion of undiagnosed infections also varies widely, ranging between 45% and 85%, highlighting gaps in national testing programs [CITATION Eur17 \l 4105]. Although our results are focused on Canada, other high-income, immigrant-receiving countries like the UK and Australia may also benefit from applying the methods developed in this paper.

A one-way sensitivity analysis identifies CHB incidence as the greatest factor influencing overall prevalence, followed by the probability of liver death in late-stage clinical health states, followed by treatment efficacy against cirrhosis. The undiagnosed fraction was sensitive to changes in diagnosis probabilities and CHB incidence, but treatment efficacy did not impact prevalence enough to have a significant effect on the undiagnosed proportion.

Limitations

The estimation method we have presented has several advantages. For one, it is systematic: the analysis we have performed can, in principle, be extended to any cohort for which data are available. Additionally, the two-step calibration method accounts for uncertainty in the model parameters that are derived from the literature, yielding statistically robust estimates. However, a few important limitations constrain its use. We made the simplifying assumption that the CHB incidence is a linear function of time, but the one-way sensitivity analysis identified CHB incidence as the parameter with the largest impact on prevalence. Therefore, the assumption of linearity greatly limits the time resolution of the model. As such, the time series presented in this paper should be taken as indicators of overall prevalence trends, rather than a detailed year-by-year accounting. In particular, the model incidence as currently formulated cannot accommodate sudden changes in vaccine uptake or immigration rates. This latter limitation may be addressed by immigrant sub-cohort modeling,

though we currently lack the data required to perform this analysis. There is also considerable uncertainty surrounding HCC incidence. For one, the HCC data we use for calibration include HCC from all causes. Following [CITATION Kim22 \l 4105], we estimate that hepatitis B accounts for 9% of HCC cases, but this estimate does not account for annual variation. Moreover, as with all cancers, HCC incidence is known to be highly age-dependent, but the available HCC data were not detailed enough to calibrate a state transition model that accounts for age. This latter limitation makes it impossible to reliably stratify our analysis by age cohort. Fortunately, both of these HCC-related limitations can be addressed with more accurate health administration data.

Conclusions

With these limitations in mind, our results point to some notable public health concerns. Despite Canada's 2016 endorsement of ambitious elimination targets, our 2021 CHB prevalence estimate does not indicate a significant decrease over previous estimates from 2013. Hepatitis B continues to be a substantial health burden for at least a hundred thousand Canadians. Of these, a significant proportion remain undiagnosed and untreated, putting them at increased risk of serious complications. Our analysis suggests that measures to decrease CHB incidence, such as vaccination and screening, have the greatest potential to decrease the overall disease burden. Based on our current data limitations, further stratified analysis is not possible. Our next step will be accessing health administrative data to overcome this barrier, with the aim of producing more reliable and detailed estimates.

We have presented a method to estimate the prevalence of CHB, using state transition model based on the natural history of the disease. The estimates are obtained via a two-stage calibration process that incorporates multiple sources of uncertainty, yielding results that are robust to perturbations of the model. This calibration process relies only on diagnosis data which are already collected by the healthcare system, making it an attractive alternative to seroprevalence studies and other methods which require the collection of novel data. Conversely, a lack of high-quality health administrative data significantly limits the accuracy of the model. This is especially true among small populations and younger age cohorts, as younger individuals will typically not have developed a large and statistically robust pattern of late-stage complications. With more data, these estimates can further be stratified by region or by age cohort, allowing them to inform the targeting of public health interventions.

Acknowledgements

This study was supported by the Public Health Agency of Canada. The authors would like to express appreciation to Dr. Curtis Cooper and Dr. Carla S. Coffin for their valuable suggestions and insights during the conduct of this study.

Conflicts of Interest

None declared.

Abbreviations

CHB: chronic hepatitis B

DC: decompensated cirrhosis

HCC: hepatocellular carcinoma

LT: liver transplant

PHAC: Public Health Agency of Canada

PLT: post-liver transplant

References

1. World Health Organization. Hepatitis B: Fact Sheet; 2024. <https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-b> [accessed 30-April-2024]
2. Government of Canada. Message from the Minister of Health - World Hepatitis Day; 2016. <https://www.canada.ca/en/health-canada/news/2016/07/message-from-the-minister-of-health-world-hepatitis-day.html> [accessed 30-April-2024]
3. Rotermann M, Langlois K, Andonov A, Trubnikov M. Seroprevalence of hepatitis B and C virus infections: Results from the 2007 to 2009 and 2009 to 2011 Canadian Health Measures Survey. Health Rep. 2013; 24(11): 3-13. PMID:24259199
4. Government of Canada. Highlights from the 2021 childhood National Immunization Coverage Survey; 2023. <https://www.canada.ca/en/public-health/services/immunization-vaccines/vaccination-coverage/2021-highlights-childhood-national-immunization-coverage-survey.html> [accessed 30-April-2024]
5. Government of Canada. Provincial and territorial routine and catch-up vaccination schedule for infants and children in Canada; 2023. <https://www.canada.ca/en/public-health/services/provincial-territorial-immunization-information/provincial-territorial-routine-vaccination-programs-infants-children.html> [accessed 30-April-2024]
6. Tran T. Management of hepatitis B in pregnancy: weighing the options. Cleveland Clinic journal of medicine. 2009; 76(Suppl 3): S25-S29. PMID:19465706
7. Statistics Canada. Table 98-10-0349-01 Place of birth and period of immigration by gender and age: Canada; 2023. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810034901>

[accessed 25-March-2025]

8. Ontario Agency for Health Protection and Promotion (Public Health Ontario). Immunization Coverage Report for School-Based Programs in Ontario: 2019-20, 2020-21 and 2021-22 School Years with Impact of Catch-up Programs. Toronto, ON: King's Printer for Ontario; 2023.
9. Biondi MJ, Marchand-Austin A, Cronin K, Nanwa N, Ravirajan V, Mandel E, et al. Prenatal hepatitis B screening, and hepatitis B burden among children, in Ontario: a descriptive study. *Cmaj*. 2020; 192(43): E1299-E1305. PMID: 33106301
10. Hamadeh A, Feng Z, Krahn M, Wong W. A model-based framework for chronic hepatitis C prevalence estimation. *PloS one*. 2019; 14(11): e0225366. doi:10.1371/journal.pone.0225366
11. Trubnikov M, Yan P, Archibald C. Hepatitis C: estimated prevalence of hepatitis C virus infection in Canada, 2011. *Canada Communicable Disease Report*. 2014; 40(19): 429. PMID:29769874
12. Longini Jr I, Byers R, Hessel N, Tan W. Estimating the stagespecific numbers of HIV infection using a Markov model and backcalculation. *Statistics in medicine*. 1992; 11(6): 831-843. doi:10.1002/sim.4780110612
13. Hamadeh A, Feng Z, Niergarth J, Wong W. Estimation of COVID-19 period prevalence and the undiagnosed population in Canadian provinces: Model-based analysis. *JMIR Public Health and Surveillance*. 2021; 7(9): e26409. PMID:34228626
14. Kim J. Model-based estimation of the prevalence of chronic hepatitis B in Canada and cost-effectiveness analysis of implementing a universal birth hepatitis B vaccination program in Ontario, Canada. University of Waterloo. 2022.
15. Terrault NA, Lok AS, McMahon BJ, Chang KM, Hwang JP, Jonas MM, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology*. 2018; 67(4): 1560-1599. PMID:29405329
16. Statistics Canada. Table 13-10-0111-01—Number and rates of new cases of primary cancer, by cancer type, age group and sex; 2024. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310011101> [accessed 30-April-2024]
17. Public Health Agency of Canada. Notifiable Diseases Online; 2024. <https://diseases.canada.ca/notifiable/> [accessed 30-April-2024]
18. Lapointe-Shaw L, Chung H, Holder L, Kwong JC, Sander B, Austin PC, et al. Diagnosis of chronic hepatitis B pericomplication: risk factors and trends over time. *Hepatology*. 2021; 73(6): 2141-2154. doi:10.1002/hep.31557
19. Wong WW, Woo G, Jenny Heathcote E, Krahn M. Cost effectiveness of screening immigrants for hepatitis B. *Liver International*. 2011; 31(8): 1179-1190. doi:10.1111/j.1478-3231.2011.02559.x
20. Hastings W. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970; 57(1): 97-109. doi:10.2307/2334940
21. Brown R, Goulder P, Matthews PC. Sexual dimorphism in chronic hepatitis B virus (HBV) infection: evidence to inform elimination efforts. *Wellcome open research*. 2022; 7: 32.

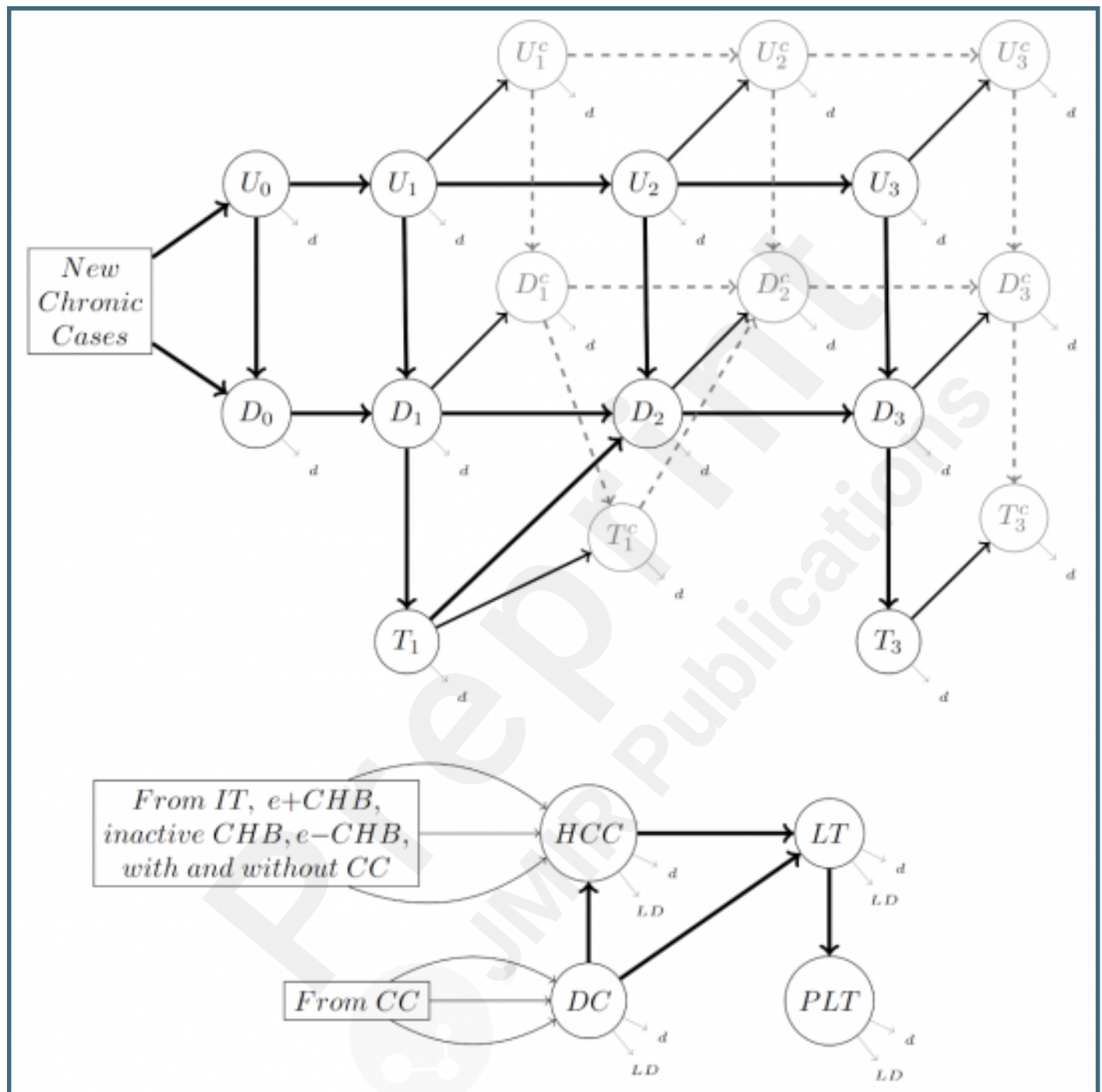
PMID: 36212217

22. Australasian Society for HIV, Viral Hepatitis and Sexual Health Medicine (ASHM). Prevalence and epidemiology of hepatitis B; 2022. <https://hepatitisb.org.au/prevalence-and-epidemiology-of-hepatitis-b/> [accessed 25-March-2025]
23. European Centre for Disease Prevention and Control. Systematic review on hepatitis B and C prevalence in the EU/EEA; 2016. <https://www.ecdc.europa.eu/en/publications-data/systematic-review-hepatitis-b-and-c-prevalence-eueea> [accessed 25-March-2025]
24. European Centre for Disease Prevention and Control. ECDC: around 9 million Europeans are affected by chronic hepatitis B or C; 2017. <https://www.ecdc.europa.eu/en/news-events/ecdc-around-9-million-europeans-are-affected-chronic-hepatitis-b-or-c> [accessed 25-March-2025]

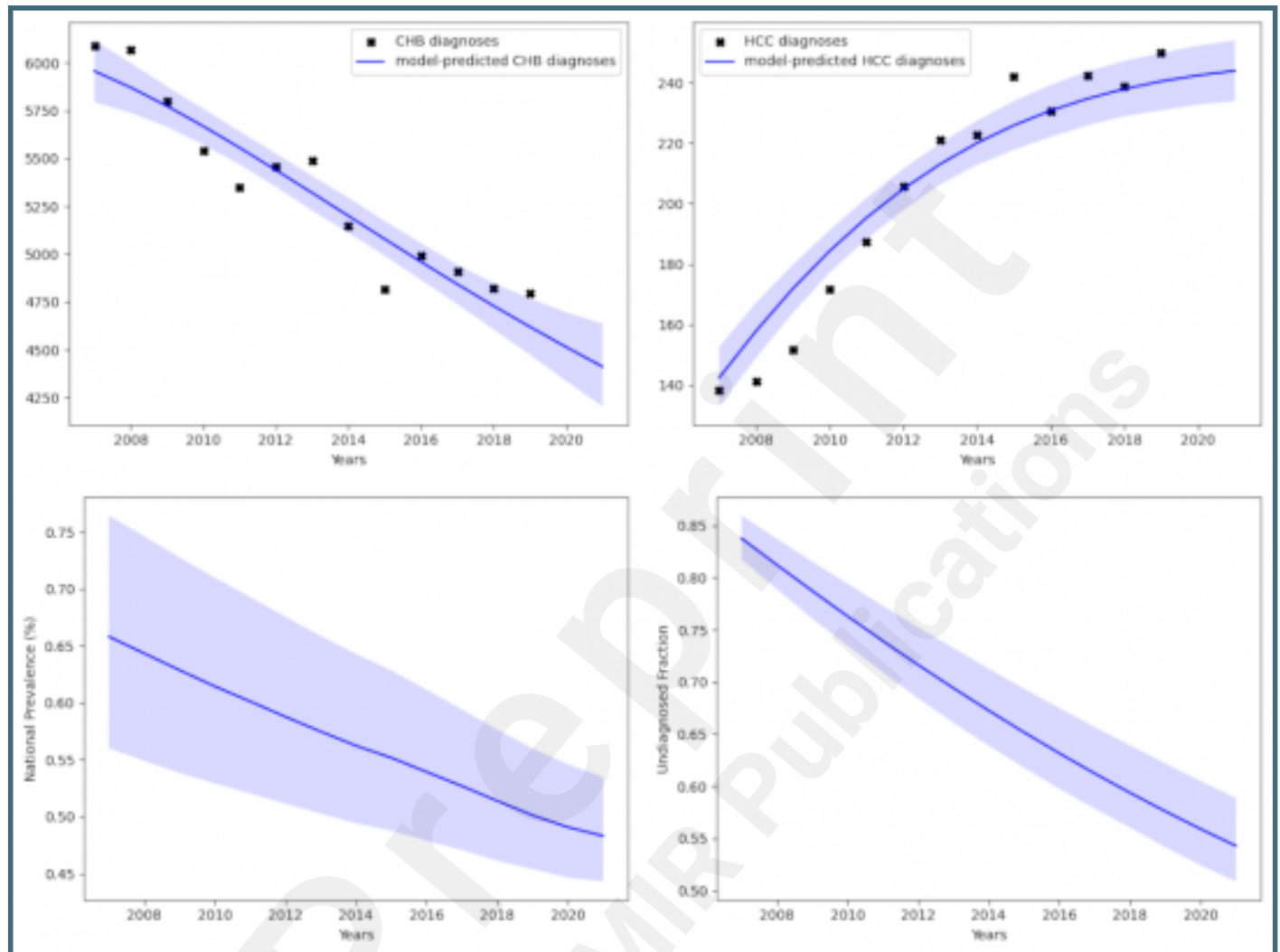
Supplementary Files

Figures

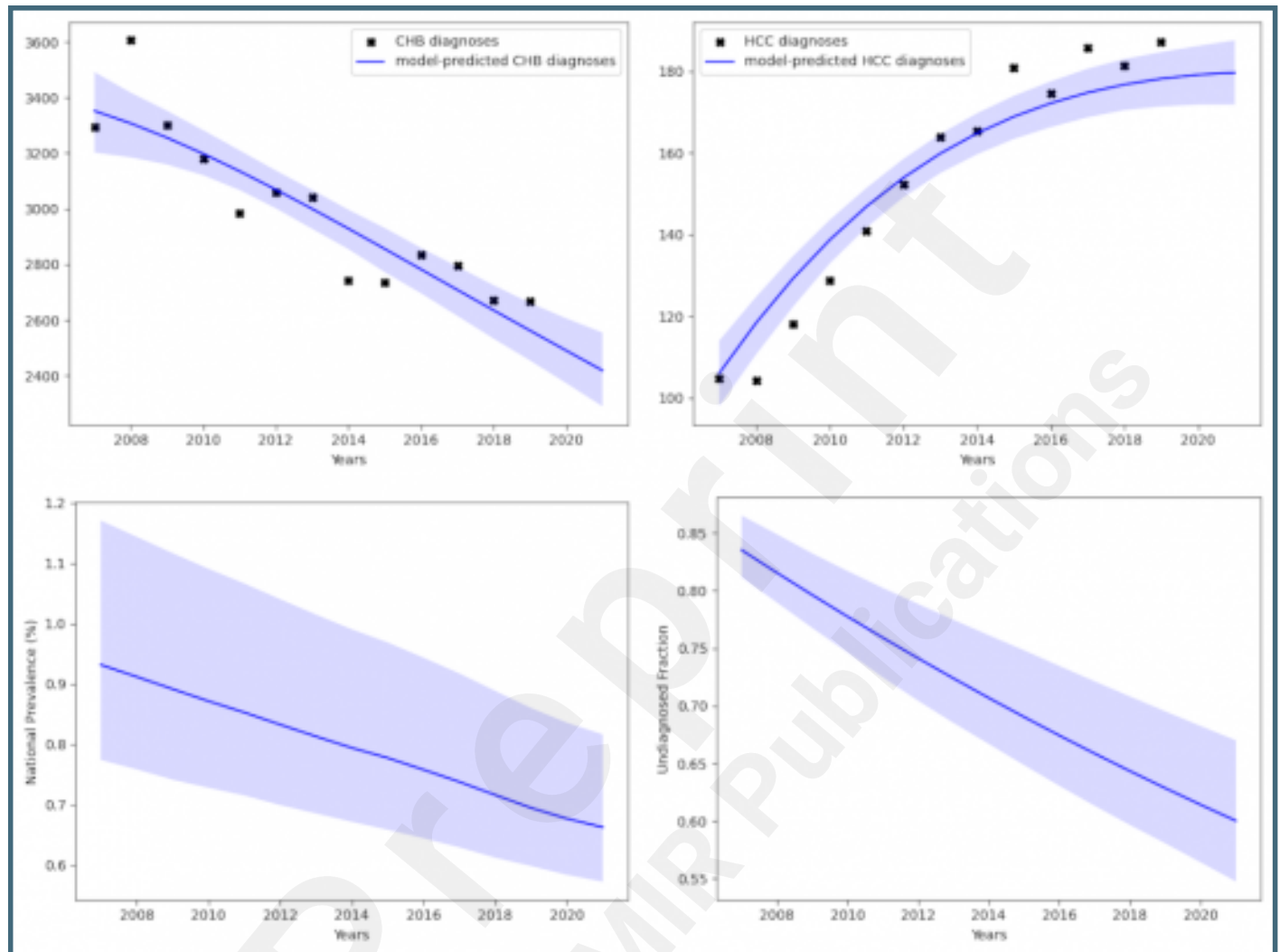
A schematic representation of the state-transition model.



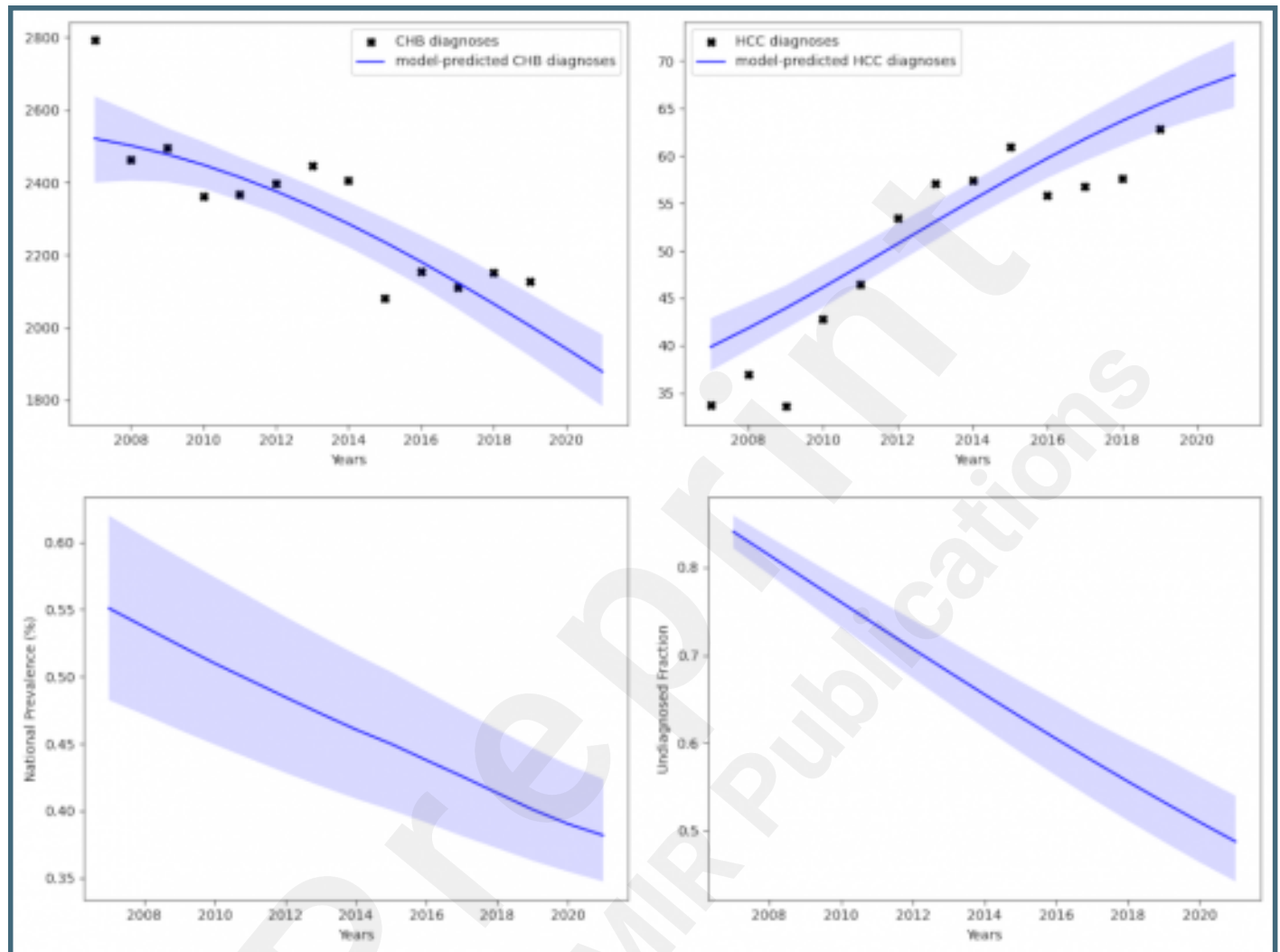
Estimates of the national CHB prevalence across all cohorts. The top two plots show the model fit compared to data for CHB diagnoses (top left) and CHB-attributed HCC diagnoses (top right). The bottom two plots accordingly display the estimates of prevalence (bottom left) and of the undiagnosed proportion (bottom right).



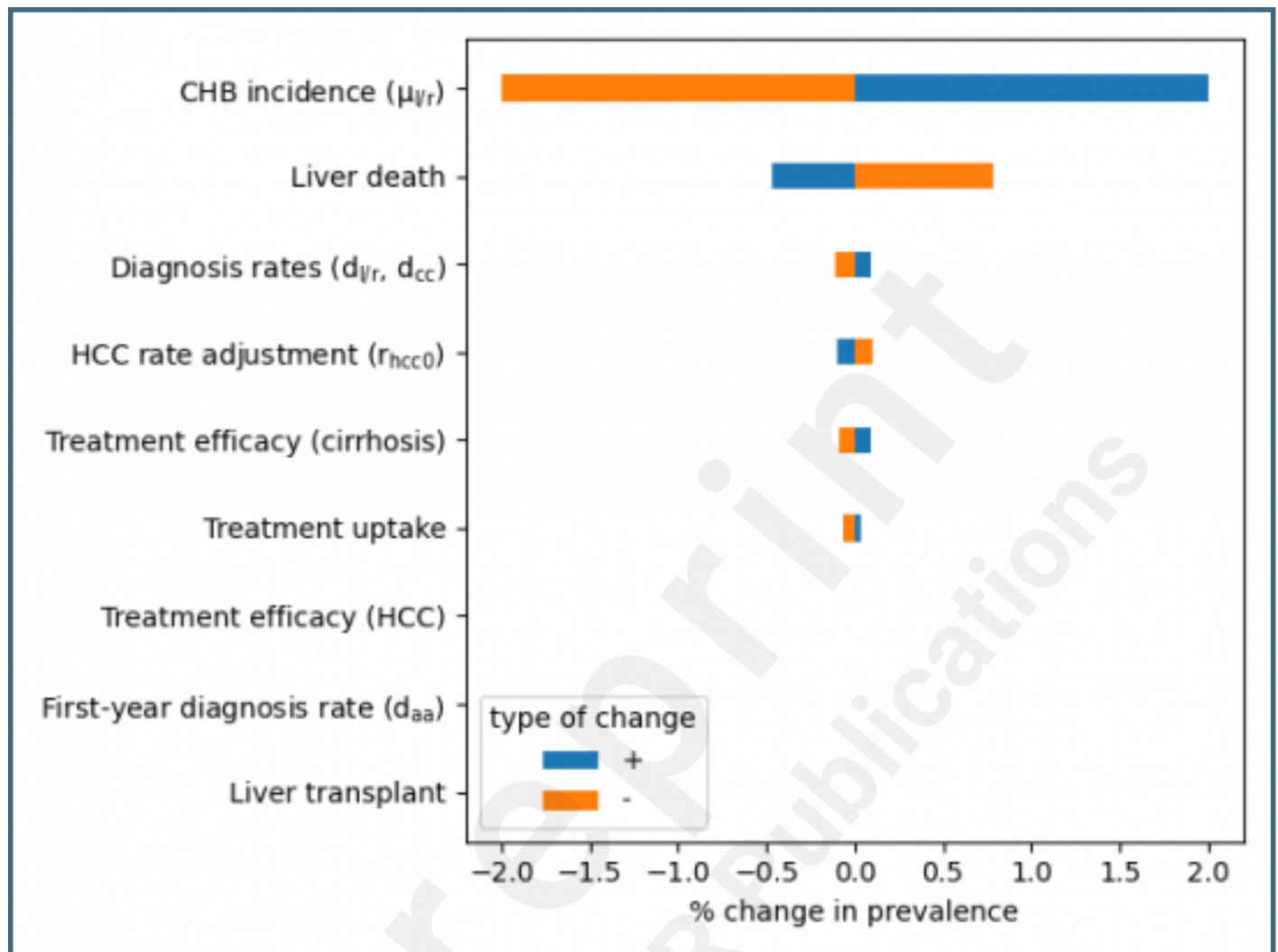
Estimates of the CHB prevalence in males. The top two plots show the model fit compared to data for CHB diagnoses (top left) and CHB-attributed HCC diagnoses (top right). The bottom two plots accordingly display the estimates of prevalence (bottom left) and of the undiagnosed proportion (bottom right).



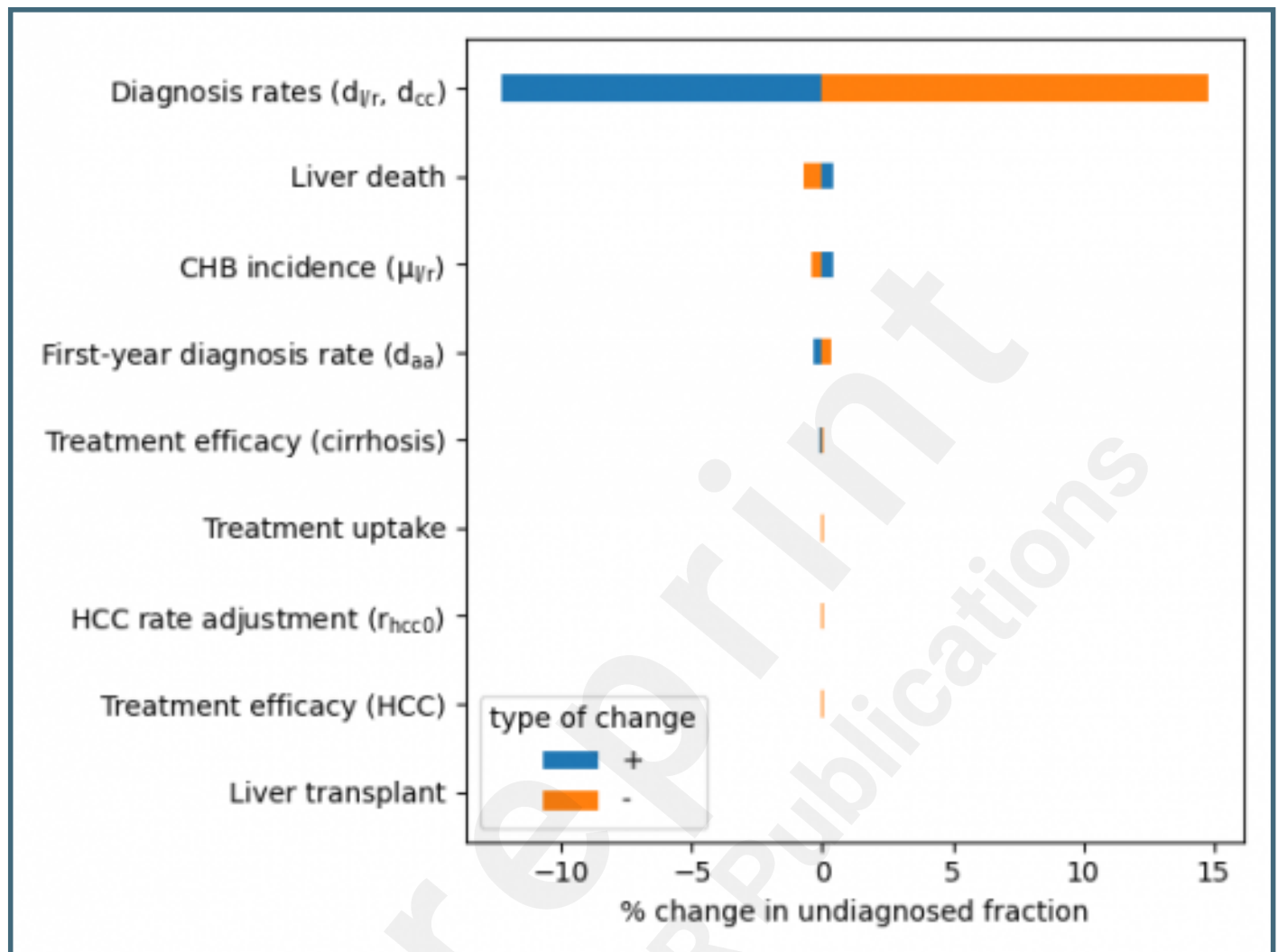
Estimates of the CHB prevalence in females. The top two plots show the model fit compared to data for CHB diagnoses (top left) and CHB-attributed HCC diagnoses (top right). The bottom two plots accordingly display the estimates of prevalence (bottom left) and of the undiagnosed proportion (bottom right).



One-way sensitivity analysis for the estimated prevalence of CHB.



One-way sensitivity analysis for the estimated undiagnosed proportion of CHB infections.



Multimedia Appendixes

Literature-derived model parameters.

URL: <http://asset.jmir.pub/assets/f51f0d83119e6ce8673de5e89c557315.docx>

Model dynamics and likelihood.

URL: <http://asset.jmir.pub/assets/e892a929ed5b63e404214a5eafcbcd66.docx>

Calibration data.

URL: <http://asset.jmir.pub/assets/6c8bde9aae45f5d90e8209f0b6b901c6.docx>

