Medical Decision
Making

**Use of a catalytic model to estimate hepatitis A incidence in
a low endemicity country: Implications for modeling
immunization policies**

Journal:	<i>Medical Decision Making</i>
Manuscript ID:	MDM-07-014.R1
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	
Complete List of Authors:	Pham, Ba'; University of Toronto, Department of Health Policy, Management and Evaluation Chen, Maggie; University of Toronto, Department of Biostatistics; University Health Network, Toronto General Hospital Tricco, Andrea; University of Ottawa, Institute of Population Health Bauch, Chris; University of Guelph, Department of Mathematics and Statistics Anonychuk, Andrea; GlaxoSmithKline, Canada, Epidemiology and Biostatistics Krahn, Murray; University of Toronto, Departments of Medicine and Pharmacy; University Health Network, Toronto General Hospital
APPLICATION AREAS:	Vaccination < GLOBAL HEALTH, PUBLIC HEALTH
DETAILED METHODOLOGY:	Systematic Reviews < EVIDENCE SYNTHESIS, MATHEMATICAL MODELS AND DECISION ANALYSIS, Population based studies < CLINICAL RESEARCH METHODOLOGY

24 Fax: 416 946 3719

25

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

¹Financial Disclosure: Financial support for this study was provided by a combined grant from the Canadian Institutes of Health Research (CIHR) and GlaxoSmithKline, Canada. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Dr. Murray Krahn is sponsored by an Investigator Award from CIHR and the F. Norman Hughes Chair in Pharmacoeconomics, Faculty of Pharmacy, University of Toronto. Andrea C. Tricco was awarded a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship and a CIHR Institute of Health Services and Policy Research Rising Star Award during the course of this study.

Abstract

Background

Evaluating the cost-effectiveness of vaccine programs with dynamic modeling requires accurate estimates of incidence over time. Since infectious diseases are often underreported, supplementary data and statistical analyses are required to estimate true incidence. We aimed to estimate the true incidence of HAV infection in Canada through a catalytic model.

Methods

Reported HAV seroprevalence and incidence data were obtained. Using a catalytic model, HAV seroprevalence data were reconciled with the corresponding cumulative risk of infection estimated from incidence data. Sensitivity analysis was conducted.

Results

The average annual reported incidence was 6.2 HAV cases per 100,000 (range 4.3-9.5) from 1980-1989 and 7.7/100,000 (5.9-10.8) from 1990-1999, indicating that Canada is a low incidence country. The seroprevalence in Canadian-born individuals (n=7 studies) was approximately 1-8% in ages <20, 1-11% in ages 20-29, 7-29% in ages 30-39, 50% in ages 40-49, and 60-82% in ages ≥50 years. Between 1980 and 1995, the catalytic model estimated an average annual incidence of 60/100,000 (95% confidence interval: 32.75, 523.83); approximately 7.73 (4.21, 67.33) times the average annual reported incidence of 7.78/100,000. For a typical birth cohort of 403,434 Canadians born in 1990, the model predicted 32,750 HAV cases by age 39, with a corresponding seroprevalence of approximately 8.12% by the year 2029.

Implications

Reliable estimates of true incidence of infectious disease are required for cost-effectiveness analysis of infectious disease programs. The catalytic model enables the synthesis of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

60 dispersed data, quantification of data limitations, and reconciliation of these limitations to
61 estimate true incidence for economic evaluations.

62 **Keywords:** Hepatitis A, prevalence, incidence, decision support techniques

63 **Word count:** 3,833 (main text), 248 (abstract), 1 Appendix, 4 Figures, 2 Tables

64 **Acknowledgements**

65 We thank Drs. Arni Rao, Bernard Duval, and Vladimir Gilca for their guidance during the
66 conduct of this research. We also thank Richard Foty for assisting with the data collection
67 and Yang Liu assisting with the data analysis.

68 **Conflict of interest**

69 At the time of study conduct, BP was a full-time employee of GlaxoSmithKline, Canada.
70 MHC, ACT, and CTB were paid to provide consulting services to GlaxoSmithKline, Canada
71 for their vaccine-related research. AMA is a full-time employee of GlaxoSmtihKline, Canada.
72 MK has nothing to declare.

73 Background

74 Hepatitis A virus (HAV) is prominent in many areas of the world (1-3). In countries with
75 low endemicity, adults are at risk for infection with HAV due to high-risk activities and
76 disease importation (4). Safe and effective HAV vaccines have been available since the mid-
77 1990s (5;6). In the United States, the hepatitis A childhood immunization strategy has
78 evolved from vaccinating children living in communities with high and medium endemicity
79 in 1996 and 1999 to routine vaccination of children nationwide in 2005 (7;8). Other low
80 endemicity countries continue with vaccine programs targeting high-risk groups (9), although
81 routine vaccination programs for children or adolescents is an alternative option.

82 In these countries, cost-effectiveness analyses of targeted versus routine vaccination
83 programs are required to inform immunization policies. The cost-effectiveness ratios
84 between targeted and routine vaccination programs are expected to be smaller than those
85 comparing routine vaccinations versus no vaccination (10). Analyses based upon dynamic
86 models of HAV transmission could increase the sensitivity of detecting such small
87 differences (11). However, knowledge of the true incidence of HAV infection is required for
88 such transmission models. Hence, robust approaches to estimate true incidence to be used as
89 input data for dynamic models of HAV transmission are necessary.

90 In most countries, the true incidence of HAV is unknown. Reported cases are an
91 unreliable indicator of the true incidence because the infection is often not clinically
92 apparent (sub-clinical) and symptomatic cases are not always reported (under-reporting) (4).
93 Measuring incidence directly with serial seroprevalence surveys is impractical because it is
94 resource-intensive (4). Data from cross-sectional seroprevalence surveys, if available, cannot
95 be used to directly infer time-specific incidence without some additional modeling steps (12).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Simple mathematical expressions were first used to model age-specific changes in the force of infection (i.e., the rate at which individuals who have not been exposed to the virus [or susceptibles] acquire infection) (13). These models were described as *catalytic* because of their structural similarities to equations commonly used in the study of chemical reactions (13). Integrating acute disease surveillance data with seroprevalence data from two national cross-sectional surveys, a catalytic model has been used to estimate the true incidence of HAV infection in the United States (12). According to results from this study, the estimate of true incidence in the United States was approximately 10 times the reported incidence and declined by 4.5% per year over the modeling duration (12). This information was subsequently used in another study to quantify the impact of hepatitis A immunization in the United States in the late 1990's (14).

The catalytic model application in the United States was based upon seroprevalence data from two national cross-sectional surveys (12). However, it is less clear whether such a catalytic model can be used to generate reliable estimates of true incidence in countries where national seroprevalence surveys are not available and alternative seroprevalence data are limited in scope and comprehensiveness. We aimed to estimate the true incidence of HAV infection in Canada, a low endemicity country with fragmented seroprevalence data.

Methods

Catalytic Model

Following the notation by Armstrong et al., the force of infection—the percentage of susceptible individuals of age a infected in year t —is given by $\lambda(a, t)$ (12). We wish to know the resulting cumulative percentage infected in a cohort that is age A in year T . The cumulative exposure $F(A, T)$ experienced by this cohort in the A years since its birth is the

force of infection experienced at age 1 in year $T-A$, plus the force of infection experienced at age 2 in year $T-A+1$, ..., plus the force of infection experienced at age a in year $T-A+a$, ..., plus the force of infection experienced at age A in year T . This cumulative exposure experienced from birth ($a=0$) to the present ($a=A$) by individuals in the birth cohort can be expressed as an integral

$$F(A, T) = \int_0^A \lambda(a, t) da \quad [1]$$

where $t = T - A + a$ from the arguments above. Since the cumulative exposure in the cohort is $F(A, T)$, the proportion of individuals in the cohort not infected by year T is given by $\exp(-F(A, T))$, hence the proportion of individuals in the cohort infected by year T is simply 1 minus that quantity, or

$$P(A, T) = 1 - \exp(-F(A, T)) = 1 - \exp\left(-\int_0^A \lambda(a, t) da\right) \quad [2]$$

To simplify estimation, one can moreover assume that age and time are independent in the force of infection; hence it can be modeled as the product of two functions:

$$\lambda(a, t) = F(a) * G(t), \quad [3]$$

where $F(a)$ is a function describing the age-specific force of infection (and is distinct from the exposure function $F(A, T)$) and $G(t)$ is a function describing how the force of infection evolves over time.

Here, the cumulative percent infected in a cohort of age A and time T , $P(A, T)$, is taken to be identical to the seroprevalence in the Canadian-born cohort of age A at time T . Thus $P(A, T)$ is hereafter referred to as seroprevalence. Similarly, the force of infection $\lambda(a, t)$ can be expressed as some function of the incidence of new infections as reflected in case reports. Therefore, equation [2] makes it possible to integrate age-stratified reported

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

incidence data with seroprevalence data in a birth cohort at a given time in order to estimate epidemiologically important parameters such as the force of infection $\lambda(a, t)$, via model fitting (14;15).

Key steps in fitting the model include: 1) obtaining seroprevalence data; 2) obtaining reported incidence data; 3) adjusting reported incidence to reflect the age-specific and time-specific components of the force of infection $\lambda(a, t)$ in equation [3], 4) fitting the catalytic model to derive parameter estimates and long-term predictions; and 5) conducting sensitivity analyses to address uncertainties in the data (15).

Step 1: Obtaining seroprevalence data

We obtained seroprevalence data from surveys of HAV antibodies in Canadian participants. The included surveys were identified through a systematic review (16). First, the seroprevalence estimates in Canadian-born cohorts that most likely reflected local transmission patterns were derived to provide data for the response variable $P(A, T)$ in equation [2] (Figure 1A). Second, the seroprevalence estimates $P_{all}(a)$ that reflected the total immunity level in the population were derived using data from all Canadians, including foreign-born individuals who might have been infected prior to immigration. The related estimates $(1 - P_{all}(a))$ denoting the proportion of susceptible individuals in age class a in the population were necessary for the adjustment in step 3 (Figure 1B).

The age-specific seroprevalence estimates $P_{all}(a)$ were derived and if necessary, pooled across survey studies using the Dersimonian-Laird random-effects meta-analytical approach (17), for age groups <10 (n=1 study) (18); 10-19 (n= 6 studies) (18-23); 20-29 (n = 6 studies) (19;21-25) 30-39 (n= 5 studies) (19;21;22;24;26); 40-49 (n = 3 studies) (21;22;26); and >50 (n = 4 studies) (19;21;25;26). Across surveys, the reported estimates $P_{all}(a)$ varied substantially with respect to studied populations, timing, and variation in the reported age

groups (19). The impact of this heterogeneity on modeled estimates and projections was assessed via sensitivity analyses (step 5).

Step 2: Obtaining reported incidence data

Data pertaining to the annual age-specific reported incidence $I(a)$ from 1980 to 2003 were obtained from the National Notifiable Disease Registry of Canada (27). According to the national surveillance program, the case definition is a laboratory confirmation of HAV infection (i.e., positive immunoglobulin M antibody to HAV) with or without symptoms. Hepatitis A incidence is known to vary cyclically, with peaks in outbreak activity every 6-8 years depending on place and time. This cyclicity is apparent in Figure 2 where one observes peaks in 1985 and 1991. Incidence appears to be climbing again after a trough in 1994, but it is clear by the year 2000 that this pattern has been interrupted. The vaccine was first introduced in targeted groups and for outbreak control in the mid 1990s (28). The main analysis was conducted using data from 1980 to 1995 where the reported incidence data were not under any influence HAV vaccination.

Step 3: Adjusting reported incidence

First, the age-specific component of the force of infection $F(a)$ in equation [2], defined as the percent of susceptible individuals of age a infected per year, was estimated as follows:

$$F(a) = \beta * I_{\text{adj}}(a) = \frac{\beta * I(a)}{(1 - P_{\text{all}}(a)) * P(J | a)}, \quad [4]$$

where the annual *reported* incidence $I(a)$ represents the percentage of individuals of age group a becoming infected per year. For each age group a , $I(a)$ was averaged from 1980-1995 to remove any short-term cyclical effects (29). Similar to the approach proposed by Armstrong and Bell approach (12), we made four adjustments to the reported incidence $I(a)$ in order to capture its contribution to the force of infection.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Adjustment 1: the reported incidence $I(a)$ was divided by $P(J|a)$, the probability of developing jaundice during acute infection in age group a , to account for sub-clinical infections which were assumed not to have been recognized or reported, thus yielding the true percentage of infections in age group a (4). Here we assumed that HAV infections without symptoms of jaundice were less likely to trigger a laboratory confirmation according to the case definition described in step 2. The estimated age-specific probabilities of jaundice $P(J|a)$ came from Armstrong and Bell (12), indicating high sub-clinical infections in children and adolescents. It was approximately 7.2% (95% confidence interval: 4.7-10.9%) in ages 0-4, 37.1% (30.7-43.8%) in ages 5-9, 70.7% (58.8-79.4%) in ages 10-17, and 85.2% (79.9-89.2%) in ages ≥ 18 .

Adjustment 2: The true percentage of infections $I(a) / P(J|a)$ was then divided by the proportion susceptible $(1 - P_{all}(a))$ since the force of infection is defined as the percentage of infections in the part of the cohort that is susceptible. This quantity is hereafter referred to as the $I_{adj}(a)$, the adjusted incidence in susceptibles;

Adjustment 3: Finally, $I_{adj}(a)$, was further adjusted by an under-reporting factor β ($\beta > 1$), representing any other sources of under-reporting in addition to what results from subclinical infection.

Adjustment 4: A common pattern that was observed in past decades in many advanced countries was a declining risk of infection due to improved sanitation and hygiene, which implies a declining force of infection (8). Hence, in equation [2], the time-specific component of the force of infection was assumed to decline exponentially from 1980 onward at an annual rate δ such that

$$G(t) = G_0 * \exp [-\delta * t], \quad [5]$$

where G_0 is a constant (12). We note that this gradual decline typically occurs over a longer timescale than the 6-8 year timescale of the natural cycling of hepatitis A incidence, which is caused by a separate phenomenon. We also note that this long-term decline is evident in the “S-shaped” age-stratified seroprevalence profile in Canadian-born individuals in Figure 1A. In contrast to the low seroprevalence in recently born cohorts, the high seroprevalence in older cohorts reflect high risk exposure in previous decades. Models that assume a constant force of infection cannot reproduce the observed S-shaped seroprevalence profiles (20;30).

Step 4: Fitting the catalytic model

Combining equations [2]-[5] yields

$$P(A, T) = 1 - \exp\left(-\int_0^A \beta * I_{adj}(a) * \exp[\delta * (M - T + A - a)] da\right), \quad [6]$$

where M denotes the mid-point of the years in which the reported incidence data were used and accounts for the constant G_0 in equation [5]. The paired parameter (β, δ) was estimated by minimizing the difference between the prevalence $P(A, T)$, which uses the seroprevalence data while the corresponding right side of equation uses the reported incidence data [6].

A smoothing function was fitted to the relationship between $I_{adj}(a)$ and age a to obtain a continuous approximation to the discrete age-structure in the reported incidence data (Appendix). The catalytic model was fitted in two steps. First, δ was varied by increments of 0.01 from 0.00001 – 1.0 (12). Given a fixed value of the parameter δ , the expression $\log(1 - P(A, T))$ according to equation [6] becomes a linear function of the parameter β . A least-square estimate of β was then derived from a univariate linear regression of $\log(1 - P(A, T)) = \beta * X$. Model goodness-of-fit was assessed using the fitted statistic R^2 and residual plots (31). Fitted values of the paired parameter $(\hat{\delta}, \hat{\beta})$ were selected to maximize the fitted

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

statistic R^2 . The 95% confidence intervals for estimates of $(\hat{\delta}, \hat{\beta})$ were derived by varying estimates of the age-specific sub-clinical infection probability $P(J | a)$ from their lower to upper 95% confidence intervals (12). Age-specific estimates of the adjusted incidence $I_{adj}(a)$ were derived according to equation [4] using the fitted values $\hat{\beta}$. Estimates of the force of infection $\lambda(a,t)$ for a hypothetical birth cohort of 1990 ($n=403,434$) were derived via equation [2] using the fitted values $(\hat{\delta}, \hat{\beta})$. Corresponding estimates of icteric or symptomatic infection incidence were calculated by applying the age-specific probability of jaundice $P(J | a)$ to estimates of the true incidence $F(a)$.

Step 5: Sensitivity analysis

In the main analysis, reported incidence data from 1980-1995 and seroprevalence data from studies with data collection in 1995 or before were used. Incidence data from 1995 to 2003 was in part influenced by targeted HAV vaccination programs, as the vaccines were first used around 1995 in Canada (5). Including this late time period, however allowed the inclusion of prevalence data from recent studies (Figure 2). In a sensitivity analysis, data from 1980 to 2003 were used.

Age-specific estimates of HAV prevalence among all Canadians varied substantially (Figure 2). In the main analysis, the pooled estimates from random-effects models for $P_{all}(a)$ were used. Sensitivity analyses were also conducted by using the lower and upper bounds of the 95% confidence intervals of the pooled estimates.

Results

The systematic search yielded 7 studies (18;19;21;22;24-26) reporting 20 point estimates for seroprevalence among Canadian-born participants (Figure 1A) (16). Seroprevalence samples

collected from these studies occurred between 1980 and 2003 (Figure 1). The seroprevalence in Canadian-born individuals was approximately 1-8% in ages <20, 1-11% in ages 20-29, 7-29% in ages 30-39, approximately 50% in ages 40-49, and 60-82% in ages ≥ 50 years (Figure 1A). A similar pattern was observed for all Canadians including foreign-born individuals, with large variation in each age group (Figure 1B). Point estimates in the age groups >40 were high and most likely influenced by the cohort effect; meaning that the high seroprevalence reflected life-long immunity from HAV infection experienced by birth cohorts in past decades (Figure 1A) (25;32).

In Canada, the average annual reported incidence was 6.2 cases per 100,000 (range 4.3-9.5) from 1980-1989, and 7.7 (5.9-0.8) from 1990-1999 (Figure 2). Coinciding with the use of HAV vaccines in targeted groups and outbreak control, there appeared to be a decreasing trend from 1995 to 2003, reaching below 2 cases per 100,000 from 2000-2003.

Modeled estimates of seroprevalence fitted the observed seroprevalence reported in survey studies for age <40 (Figure 3). Limited seroprevalence data were available for modeled estimates in ages ≥ 40 (n=3 data points, Figure 3). The annual rate of decline in the force of infection over time was estimated to be $\delta = 0.12$ (95% confidence interval: 0.08, 0.15; Table 1). The under-reporting factor estimated from the linear regression involving the response $\log(1-P(A,T))$ derived from equation [6] was $\beta = 7.73$ (4.21, 67.33), with a goodness-of-fit statistic $R^2 = 0.86$. This under-reporting estimate suggests that for every reported HAV case, approximately 8 cases were not reported. In the sensitivity analysis, estimates of the annual declining rate δ and under-reporting factor β exhibited large confidence intervals. The maximum boundaries across different confidence intervals for δ ranged from 0.05 to 0.15, and for β from 4.21 to 76.83 (Table 1).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

281 The reported incidence was adjusted to better reflect estimates of the force of
282 infection (Figure 4). In ages 0 to 9, reported incidence substantially under-estimated the true
283 incidence, most possibly due to sub-clinical infection. For age groups ≥ 30 , reported
284 incidence also under-estimated true incidence. Note that the seroprevalence estimates among
285 all Canadians increased with age, leading to a decrease in the relative size of susceptibles in
286 the older age groups and an increase in the gap between reported and adjusted incidence
287 according to equation [4].

288 The model estimated an average annual incidence of 60.14 HAV cases per 100,000 in
289 Canada (95% confidence interval: 32.75, 523.83); approximately 7.73 times the average
290 annual reported incidence of 7.78 cases per 100,000 from 1980 to 1995. For a typical birth
291 cohort of 403,434 Canadians born in 1990, the model predicted 32,750 HAV cases by age
292 39, with a corresponding prevalence of approximately 8.12% by the year 2029 (Table 2). The
293 cumulative number of cases in older age groups was predicted taking both the cohort effect
294 and susceptibility into account (Table 2).

295

296 **Discussion**

297 This paper provides an example of how catalytic modeling can be employed to bridge the
298 gap between reported incidence and seroprevalence data in order to estimate the true
299 incidence pattern over time in a population. This information can be used to understand
300 transmission dynamics of infectious diseases over time by distinguishing age and cohort
301 effects on the force of infection. This type of modeling is particularly salient for
302 epidemiologists and economic modelers attempting to estimate the future effects of
303 infectious disease control programs, particularly vaccines, on a population's health.

Cost-effectiveness analyses of disease control programs often rely on cohort models, which essentially consider the individual in isolation from the population and do not take population-level mechanisms of disease transmission into account (10). They implicitly assume that the force of infection does not change when control programs are implemented, which is rarely the case. Hence, they can either under-estimate or over-estimate cost-effectiveness of the intervention (11). In contrast, dynamic models take into account mechanisms that govern the time evolution of epidemiological measures, such as the number of infected individuals, mortality rates, and force of infection. In principle, they are more accurate than corresponding cohort models for cost-effectiveness analysis of control programs (10;11). However, application of dynamic models requires knowledge of the true incidence of infection; hence catalytic modeling or a similar approach is necessary.

The catalytic model described in this research was first proposed by Armstrong and Bell (12), where they reconciled the reported incidence of HAV in the United States with the observed prevalence of HAV antibody from two nation-wide surveys (12). In the current application of the model, fragmented and heterogeneous seroprevalence data identified through a systematic view was used (16). The data was of low quality (16), substantially contributing to large confidence intervals around key parameter estimates. Consequently, the model was not robust as it is constrained by the lack of reliable data.

An advantage of catalytic modeling is that it allows the identification of cohort effects (i.e., changes in the force of infection on generational timescales) (2). Cohort effects are reflected in S-shaped age-structured seroprevalence profiles (e.g., Figure 2). Age-specific data from a prevalence survey at one point in time can not distinguish between age- and time-dependence (33). Data from multiple studies are under the influence of both time-specific and age-specific effects. Note that HAV transmission depends on hygiene, food

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

328 preparation, or contact with infective individuals. These factors change over time and risk-
329 related behaviors could change across generations. Estimates of true incidence are biased
330 unless both age- and time-specific effects in seroprevalence data are modeled (33). The
331 current model fulfilled this requirement. Failure to account for cohort effects has wide
332 implications for interpreting seroprevalence data and predicting the impact of vaccination
333 programs with dynamic models (30).

334 Simple catalytic models have been used previously to describe the age-specific
335 prevalence of susceptibility to rubella virus (34;35), to estimate HAV prevalence rates over
336 times as part of an assessment of the temporal relationship between childhood acute
337 lymphoblastic leukemia and hygiene conditions (36), and to estimate true incidence of
338 primary toxoplasmosis infection (33), HCV infection (37), and HSV-2 infection (38).
339 Catalytic modeling was also used to estimate age-related rates of measles infection (14), and
340 to study the age distribution of attack of measles (29).

341 Early catalytic models focused on the functional form of the age-specific force of
342 infection $\lambda(a)$ (13). Other functional forms have been used: constant since birth (39),
343 constant over the range of data, but not necessarily since birth (13;29;40); higher order
344 polynomials (14;41); and an exponentially damped linear function (42). Nonparametric
345 functions have also been introduced (43). Ades and Nokes suggest that flexible functional
346 forms might have the disadvantage of occasionally generating extreme curvatures at the
347 edges of the data, precisely where prediction of future trends is desired (33). Griffiths
348 provided an approach to graphically presenting observed age distributions to reveal the
349 nature of dependence of the true incidence upon age (29).

350 In the current model, the time-dependent component of the force of infection (i.e.,
351 $G(t)$) was modeled using an exponential function to allow for the cohort effect. However,

recent research suggests that the temporal evolution of the force of infection in Canada in the past century can be well-described by a sigmoidal function, where transmissibility declines from a high constant level to a low constant level with a relatively rapid period of transition in between (30). The period of rapid transition may correspond to significant post-war improvements in sanitation and hygiene, for example. The use of an exponential function may partly explain the shape of the curve in Figure 3, where seroprevalence appears to be increasing exponentially with age in the model but not in the data. A sigmoidal function might produce a better fit to the data, and because it is not piecewise, yields an easier analysis. Future research may explore the use of a sigmoidal function in catalytic modeling.

In addition to offering an example of the use of catalytic modeling with fragmented and limited data, our results also highlight the potential biases associated with the use of readily-available epidemiological data in cost-effectiveness evaluation of infectious disease control programs. Data from notifiable disease reporting substantially under-estimates true incidence due to sub-clinical symptoms in HAV infection (4) and under-reporting of clinical cases (12). Prevalence data from cross-sectional surveys was fragmented and suffered from a cohort effect; the high prevalence in the older age groups most likely reflected infection risk of past decades (20;30). Reported prevalence from surveys including foreign-born subjects might not exclusively reflect local transmissibility. The uncertainty induced by these potential sources of bias in the input data should be recognized, quantified, and assessed in cost-effectiveness analyses of infectious disease control programs.

A number of key assumptions were made in the catalytic model we used. First, we assumed that HAV infection leads to life-long immunity. The age-dependent force of infection was assumed to be constant and the susceptible population homogeneous with

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

376 respect to infection exposure. A population with stable age distribution was assumed, as it
377 was convenient to ignore deaths and migration to and from the population (29). With
378 respect to input data to the model, seroprevalence estimates from small studies and
379 heterogeneous study populations were generalized to represent Canadian estimates.

380 In summary, catalytic models integrate reported incidence and seroprevalence data
381 providing reasonably reliable estimates of past, current, and projected force of infection.
382 They provide a logical structure to link highly dispersed data, quantify limitations in the data
383 sources, and reconcile these limitations to derive estimates of true incidence. They are useful
384 to epidemiologic and economic modelers attempting to estimate the cost-effectiveness of
385 disease control programs through dynamic modeling.

Appendix

A 2-parameter exponential function form was used for the relationship between $I_{adj}(a)$ and age a , $I_{adj}(a) = \exp(-\alpha * a^\tau)$. Fitted values for (α, τ) were obtained through fitting the univariate linear regression $\log(-\log(I_{adj}(a))) = \alpha + \tau * \log(a)$. This resulted in $\alpha = 5.40$ (95% confidence interval: 5.10, 5.87), $\tau = 0.15$ (0.13, 0.18), and a goodness-of-fit statistic R^2 of 0.92.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Panel 1A: Prevalence among Canadian-born study participants Panel 1B: Prevalence among all Canadians, including foreign-born participants

Figure 1: Age-specific prevalence (%) abstracted from clinical studies identified through a systematic review (16)

Notes: The column “Author” displays the first author and year of publication of the included studies; “Population” the major age groups, participant characteristics, and year of data collection in the included studies; “Age” the age groups according to the original study reports; “n” sample size. Plotted lines are percent estimates of participants with positive HAV antibodies and their corresponding 95% confidence intervals. A total of 7 studies contributed 20 data points in Panel 1A and 10 studies contributed 30 data points to Panel 1B.

Abbreviations: NS Nova Scotia, BC British Columbia, G6 Students Grade 6 students, and CF Canadian Forces.

Figure 2: Age-adjusted reported incidence of hepatitis A in Canada from 1980 to 2003

Figure 3: Observed seroprevalence reported in seven clinical studies identified from the systematic review (i.e., 20 data points are represented by the dots) and corresponding predicted prevalence denoted by the fitted line.

Figure 4: Reported and adjusted average annual incidence of hepatitis A in Canada, 1980-1995.

Notes: Reported incidence data was available for each decade of age except for the age group 40-59 and 60+. The blue bars show the unadjusted reported incidence $I(a)$. The red bars show incidence after adjustment for sub-clinical infection using the age-specific probability of jaundice $P(J | a)$. The yellow bars show incidence after further adjusting the population denominator for the relative size of susceptibles $[I(a)/(1-P_{all}(a))]$. Additional adjustment for under-reporting was possible according to expression [3] (data not shown in the graph).

Sensitivity Analyses	Declining Trend Parameter δ (95% CI)	Under-Reporting Factor β (95% CI)
Varied incidence data		
Main analysis with 1980-1995 data	0.12 (0.08, 0.15)	7.73 (4.21, 67.33)
Sensitivity analysis with 1980-2003 data	0.10 (0.07, 0.15)	9.83 (6.39, 63.65)
Varied HAV prevalence in all Canadians		
Main analysis with pooled estimates ¹	0.12 (0.08, 0.15)	7.73 (4.21, 67.33)
Sensitivity analysis with lower bound ²	0.09 (0.07, 0.13)	16.56 (6.89, 76.83)
Sensitivity analysis with upper bound ³	0.13 (0.05, 0.14)	3.88 (2.69, 48.23)

Table 1: Parameter estimates from the main analysis and corresponding sensitivity analyses. **Notes:** ¹Pooled estimates from random-effects model of HAV prevalence in all Canadians (data from Figure 2 – Panel B) were used to adjust reported incidence data for the proportion susceptibles (i.e., in expression [3]). ²Lower bounds from 95% confidence intervals of the pooled estimates from footnote 1. ³Upper bounds from 95% confidence intervals of the pooled estimates from footnote 1. **Abbreviation:** CI confidence interval.

Age Group	Year	HAV cases*	Icteric cases*	HAV cases / 100,000	Icteric cases / 100,000	Estimated Prevalence (%)
0-9	1990-1999	7,261	362	1,800	90	1.80
10-19	2000-2009	3,954	492	980	122	2.78
20-29	2010-2019	12,346	599	3,060	148	5.84
30 - 39	2020-2029	9,198	238	2,280	59	8.12

Table 2: Long-term projection from the catalytic model for true incidence of HAV infection in a Canadian birth cohort of 1990 (n=403,434).

Note: *Cumulative number of cases in each age group (e.g., 7,261 cases from birth to age 9 in the first age group).

Reference List

(1) Cuthbert JA. Hepatitis A: old and new. Clin Microbiol Rev 2001 Jan;14(1):38-58.

(2) Gust ID. Epidemiological patterns of hepatitis A in different parts of the world. Vaccine 1992;10 Suppl 1:S56-S58.

(3) Tapia-Conyer R, Santos JI, Cavalcanti AM, Urdaneta E, Rivera L, Manterola A, et al. Hepatitis A in Latin America: a changing epidemiologic pattern. Am J Trop Med Hyg 1999 Nov;61(5):825-9.

(4) Scheifele DW. Hepatitis A vaccines: the growing case for universal immunisation of children. Expert Opin Pharmacother 2005 Feb;6(2):157-64.

(5) Scheifele DW, Ochnio J. Hepatitis A vaccine: is it being used to best advantage? CMAJ 2002 Jul 9;167(1):44-5.

(6) Furesz J, Scheifele DW, Palkonyay L. Safety and effectiveness of the new inactivated hepatitis A virus vaccine. CMAJ 1995 Feb 1;152(3):343-8.

(7) Fiore AE, Wasley A, Bell BP. Prevention of hepatitis A through active or passive immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2006 May 19;55(RR-7):1-23.

(8) Samandari T, Bell BP, Armstrong GL. Quantifying the impact of hepatitis A immunization in the United States, 1995-2001. Vaccine 2004 Oct 22;22(31-32):4342-50.

(9) Stene-Johansen K, Tjon G, Schreier E, Bremer V, Bruisten S, Ngui SL, et al. Molecular epidemiological studies show that hepatitis A virus is endemic among active homosexual men in Europe. J Med Virol 2007 Apr;79(4):356-65.

(10) Crewe MD, Embil JA, Garner JB. Prevalence of antibodies to hepatitis A virus in Nova Scotia children. Can Med Assoc J 1983 May 15;128(10):1195-7.

(11) Edmunds WJ, Medley GF, Nokes DJ. Evaluating the cost-effectiveness of vaccination programmes: a dynamic perspective. Stat Med 1999 Dec 15;18(23):3263-82.

(12) Armstrong GL, Bell BP. Hepatitis A virus infections in the United States: model-based estimates and implications for childhood immunization. Pediatrics 2002 May;109(5):839-45.

(13) Muench H. Catalytic Models in Epidemiology. Cambridge, MA: Harvard University Press; 1959.

- (14) Grenfell BT, Anderson RM. The estimation of age-related rates of infection from case notifications and serological data. *J Hyg (Lond)* 1985 Oct;95(2):419-36.
- (15) Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics* 2006;24(4):355-71.
- (16) Pham B, Duval B, De SG, Gilca V, Tricco AC, Ochnio J, et al. Seroprevalence of hepatitis A infection in a low endemicity country: a systematic review. *BMC Infect Dis* 2005;5:56.
- (17) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986 Sep;7(3):177-88.
- (18) Duval B, De Serres G, Gilca V, Ochnio J, Scheifele DW. Nationwide Canadian study of hepatitis A antibody prevalence among children eight to thirteen years old. *Pediatr Infect Dis J* 2005;24(6):514-9.
- (19) Levy I, Chen D, Sherman M, Smith D, Krajden M. Hepatitis A virus seroprevalence in 1,000 university students in Toronto. *Can Commun Dis Rep* 2001 Jun 1;27(11):93-6.
- (20) Ochnio JJ, Scheifele DW, Ho M. Hepatitis A virus infections in urban children--are preventive opportunities being missed? *J Infect Dis* 1997 Dec;176(6):1610-3.
- (21) Buchner BK, Staszak M, Duravetz J, Moore BP. Hepatitis markers detected in a Canadian blood donor population. *Acta Haematol Pol* 1980 Apr;11(2):69-71.
- (22) Cook D, Wilton L, Patrick D, Zou S, Sherman M, Krajden M. Prevalence of antibodies to hepatitis A virus in a cohort of women of child-bearing age. 68th Canadian Association for Clinical Microbiology and Infectious Disease, Nov 5-8, 2000 - Ottawa . 11-5-2000.
Ref Type: Generic
- (23) Embil JA, Manley K, White LA. Hepatitis A: a serological study in the Canadian Armed Forces. *Mil Med* 1989 Sep;154(9):461-5.
- (24) Ochnio JJ, Scheifele DW, Marion SA, Bigham M, Ho M. Hepatitis A virus (HAV) infections in young adults in British Columbia: feasibility of a novel mail-based approach to population sampling. *Proceedings of the 7th Annual Conference on Vaccine Research. Arlington, VA, USA Abstract P4. 1-1-2004.*
Ref Type: Abstract
- (25) McFarlane ES, Embil JA, Manuel FR, Gorelick M. Prevalence of antibodies to hepatitis A antigen in patients attending a clinic for treatment of sexually transmitted diseases. *Sex Transm Dis* 1980 Apr;7(2):87-9.

(26) De Serres G, Levesque B, Higgins R, Major M, Laliberte D, Boulianne N, et al. Need for vaccination of sewer workers against leptospirosis and hepatitis A. *Occup Environ Med* 1995;52(8):505-7.

(27) Public Health Agency of Canada. Notifiable Diseases Online. http://dsol-smed.phac-aspc.gc.ca/dsol-smed/ndis/index_e.html 2005 April 5

(28) Wu J, Zou S, Giulivi A. Current hepatitis A status in Canada. *Can J Infect Dis Med Microbiol* 2001;12(6):341-4.

(29) Griffiths DA. A catalytic model of infection for measles. *Applied Statistics* 1974;23:330-9.

(30) Srinivasa Rao AS, Chen MH, Pham BZ, Tricco AC, Gilca V, Duval B, et al. Cohort effects in dynamic models and their impact on vaccination programmes: an example from Hepatitis A. *BMC Infect Dis* 2006;6:174.

(31) Christensen R. Plane questions to complex questions: The theory of linear models. New York: Springer-Verlag; 1987.

(32) Payment P. Antibody Levels to Selected Enteric Viruses in a French-Canadian Population in the Province of Quebec. *Immunol Infect Dis* 1991;1:317-22.

(33) Ades AE, Nokes DJ. Modeling age- and time-specific incidence from seroprevalence: toxoplasmosis. *Am J Epidemiol* 1993 May 1;137(9):1022-34.

(34) Cutts FT, Vynnycky E. Modelling the incidence of congenital rubella syndrome in developing countries. *Int J Epidemiol* 1999 Dec;28(6):1176-84.

(35) Cutts FT, Abebe A, Messele T, Dejene A, Enquselassie F, Nigatu W, et al. Sero-epidemiology of rubella in the urban population of Addis Ababa, Ethiopia. *Epidemiol Infect* 2000 Jun;124(3):467-79.

(36) Smith MA, Simon R, Strickler HD, McQuillan G, Ries LA, Linet MS. Evidence that childhood acute lymphoblastic leukemia is associated with an infectious agent linked to hygiene conditions. *Cancer Causes Control* 1998 May;9(3):285-98.

(37) Armstrong GL, Alter MJ, McQuillan GM, Margolis HS. The past incidence of hepatitis C virus infection: implications for the future burden of chronic liver disease in the United States. *Hepatology* 2000 Mar;31(3):777-82.

(38) Armstrong GL, Schillinger J, Markowitz L, Nahmias AJ, Johnson RE, McQuillan GM, et al. Incidence of herpes simplex virus type 2 infection in the United States. *Am J Epidemiol* 2001 May 1;153(9):912-20.

(39) Beutels M, Van DP, Aelvoet W, Desmyter J, Dondeyne F, Goilav C, et al. Prevalence of hepatitis A, B and C in the Flemish population. *Eur J Epidemiol* 1997 Apr;13(3):275-80.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- (40) Thiry N, Beutels P, Shkedy Z, Vranckx R, Vandermeulen C, Wielen MV, et al. The seroepidemiology of primary varicella-zoster virus infection in Flanders (Belgium). *Eur J Pediatr* 2002 Nov;161(11):588-93.
- (41) Farrington CP, Kanaan MN, Gay NJ. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data (with discussion). *Applied Statistics* 2001;50:251-92.
- (42) Edmunds WJ, Gay NJ, Kretzschmar M, Pebody RG, Wachmann H. The pre-vaccination epidemiology of measles, mumps and rubella in Europe: implications for modelling studies. *Epidemiol Infect* 2000 Dec;125(3):635-50.
- (43) Shkedy Z, Aerts M, Molenberghs G, Beutels P, Van DP. Modelling age-dependent force of infection from prevalence data using fractional polynomials. *Stat Med* 2006 May 15;25(9):1577-91.