A Simulation Model for Major Depression Deriving from the Canadian Community Health Survey, Mental Health and Wellbeing (CCHS 1.2)

Scott B. Patten MD, PhD.*

* Professor, Departments of Community Health Sciences and Psychiatry, University of Calgary.

Part 1. Simulation Strategy

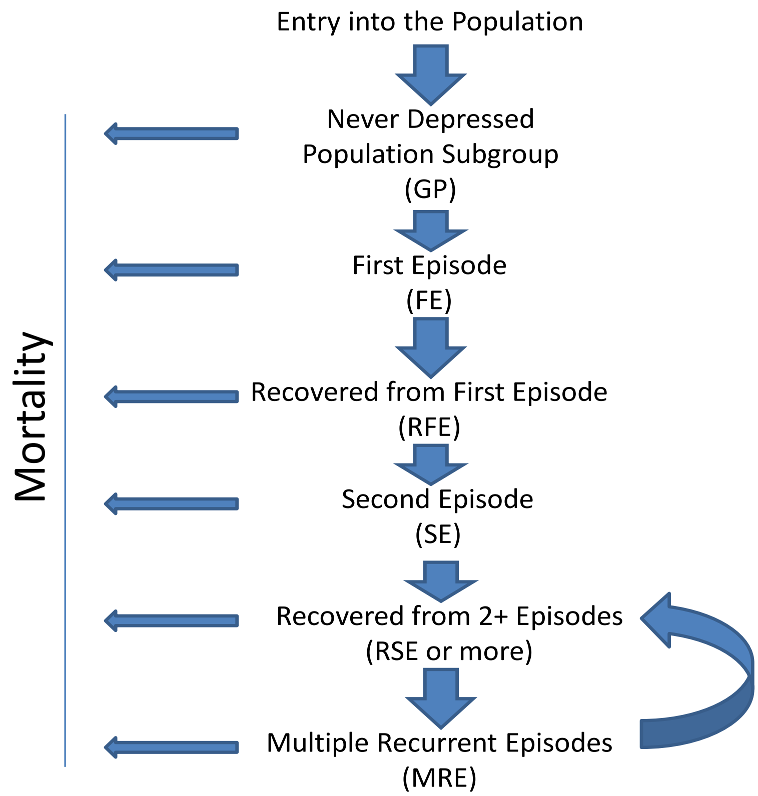
Background

This goal of this project was to explore opportunities for population-based planning for mental health interventions using simulation models. Mindfulness Based Cognitive Behavioural Therapy (MBCT) was adopted as a specific example of the sort of situation in which simulation may be helpful. Existing literature suggests that the therapy may be effective at reducing relapse rates, but only in people that have had three or more past episodes of major depression. The proportion of the population with a history of three or more past episodes can be directly estimated from available survey data. However, another key parameter is the rate at which people with three or more episodes emerge within the population. This latter parameter would help to determine the rate at which an identifiable need for this therapy emerges within the population. However, such a parameter cannot be directly estimated from cross-sectional survey data. Ideally, the estimate would be available from a large-scale longitudinal study. However, no suitable data sources exist in Canada. Another option would be simply to "make a guess." However, the underlying epidemiology is sufficiently complex that such guesses could easily be wrong. A simulation model can bring an added degree of quantification and the can support the integration of available information to assist with the goal of quantification. A simulation model produces a tool that can be refined when new data become available and that can be used to evaluate "what if" scenarios to assist with policy formation.

The best data source for mental health epidemiology in Canada is a cross-sectional survey called the Canadian Community Health Survey (CCHS), iteration 1.2, Mental Health and Wellbeing (CCHS 1.2). Details of the survey may be found on the Statistics Canada web page (www.statcan.gc.ca), and in a report by Gravel & Béland (1). The CCHS 1.2 is a cross-sectional survey similar to those that have now been conducted in many countries around the world. The survey included a Canadian adaptation of the Composite International Diagnostic Interview and collected information on a variety of parameters relevant to the longitudinal course of mental disorders in the population. The existence of this database makes it possible to develop a simulation models that reflect the longitudinal epidemiology more fully than traditional targets of epidemiologic information such as prevalence proportions or odds ratios.

In this project, the general approach to simulation was to develop an incidence-prevalence model, depicting prevalence as a steady-state pool within the population. The prevalence pool represents the segment of the population who are depressed at a point in time. There is an inflow into this prevalence pool from incidence and recurrence. In turn, the prevalence pool is drained by recovery and mortality. This paradigm is widely used in epidemiology because the prevalence of an endemic disease can be conceptualized as a steady state balance between inflow and outflow. A schematic description of the approach is presented in Figure 1.

Figure 1.*



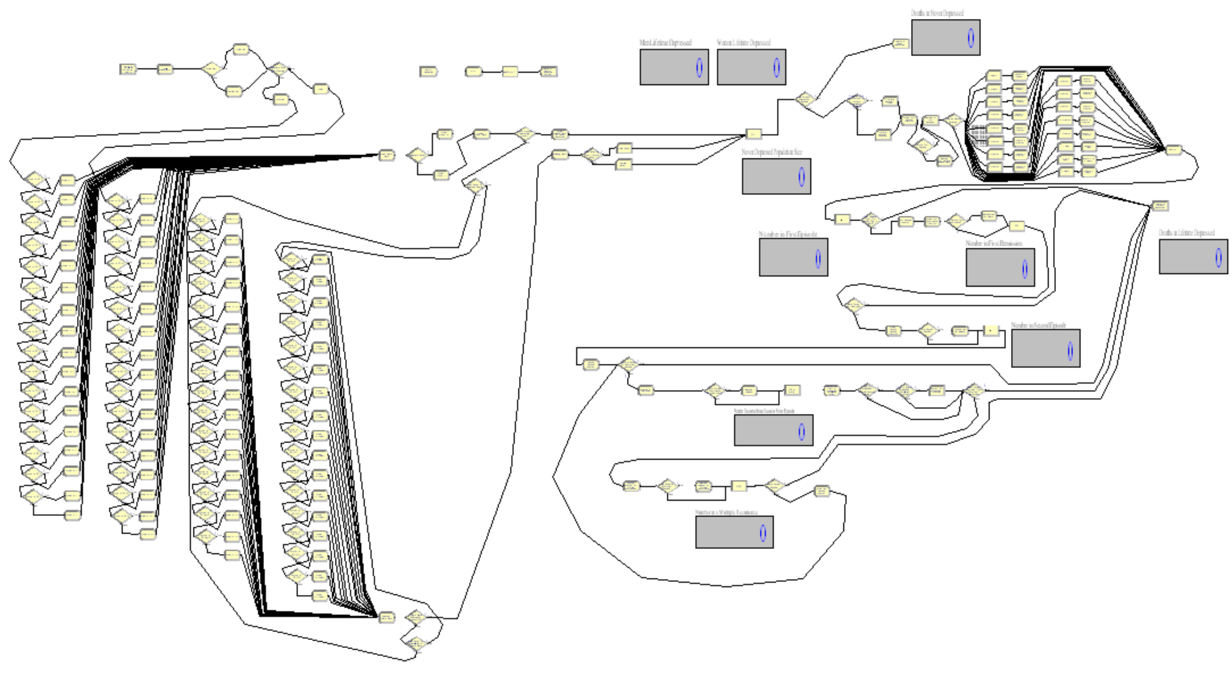
* the abbreviation "GP" represents the general population not yet having had a major depressive episode (or who will never have one), other abbreviations are first episode (FE), recovered from first episode (RFE), second episode (SE), recovered from second or more episodes (RSE or more) and multiple recurrent episodes (MRE).

The simulation model was developed in the SIMAN language using the software Arena®(2). Arena® provides a graphic interface for model development. The modeling uses discrete event simulation, an approach centered on simulation of time to occurrence of events. A screen capture of the entire model is presented in Figure 2. The Figure depicts icons of various shapes joined by lines. The icons represent modules within the model, some of which constitute intrinsic aspects of the simulation model and others that are used to record output from the model. Simulation output needs to be recorded so that it can be compared to epidemiologic data for purposes of model calibration. The lines connecting the icons are the paths followed by model "entities" – which in the model presented here are simulated people. As they represent people with different characteristics, for example, those with and without depression, each entity is characterized by attributes. These attributes are variables within the model that are attached to specific entities. For example, an attribute representing whether or not a person has current major depression can be depicted by an attribute variable assuming values of 0 (not in a major depressive episode) and 1 (is in a major depressive episode). Each entity also has an attribute reflecting whether the entity is male or female. Figure 2 is very complex because all of the modules in the model are depicted in the screen capture. In subsequent sections of this document the model is broken down into parts, each of which is described in more detail.

As noted, the icons in the screen capture (Figure 2) represent modules performing various functions within the model or recording model output. One type of module is a "generate" module. This module produces model entities (representing individual people within the simulation). In the model presented here, entities enter the simulation at age 15. Time is actually depicted in days, but an attribute is added at the time an entity enters the model representing 15 years of age (age in days = age in years multiplied by 365.2 days per year). This decision was made because age 15 was the lower limit of the sampling frame for CCHS 1.2. Another type of module is a "decide" module. These modules determine the path that an entity follows through the model either on a probabilistic basis or by channeling entities depending on their attributes. This type of module is used, for example, in order to allow the simulation of mortality to be sex specific. An entity enters the model from the "Generate" module, following the model path. Later, an "assign" modules assigns an attribute for sex. Still later in the model, in a section designed to simulate mortality, a "decide" module channels the entities down male or female paths conditional on their sex attribute. Both male and female paths lead to additional "decide" modules which simulate the mortality experience probabilistically using age and sex specific mortality rates from vital statistics data, in turn deriving from Canada's national statistical agency, Statistics Canada. Because the "decide" module channels male and female attributes into different paths, the model can simulate their mortality using different sets of mortality rates (in this case, sets of age-specific rates) for men and women. The process of simulation is described in more detail below and examples are mentioned briefly here only to illustrate the role of the different types of modules. "Record" modules save attribute values for later processing of the model output. Since this is an epidemiologic model, most of the "record" modules count the number of entities having specific attributes. These counts are used by the model to calculate variables representing its outputs, which are depictions of epidemiologic parameters that can be estimated in the CCHS 1.2. For example, by counting the number of entities with and without the major depressive episode attribute at a point in time, the model can simulate point prevalence. Similarly, by counting entities having an attribute for "ever depressed" the model can simulate lifetime prevalence. The end of each entity's participation in the simulation occurs when a "dispose" module is entered. In the current model, this disposal represents the death of that entity. The final type of module used in the model is a "queue." In discrete event simulation, the simulation model is based on simulating time to the occurrence of events. When an entity becomes depressed (ie. the attribute representing depression is set by the model to a value that represents depression) another attribute representing the simulated duration of that episode is assigned to the entity. The entity's status as a depressed person is represented by its residing in a "queue" for one of the depressed states (called FE, SE & MRE, see Figure 1). In these instances, the "queue" module is programmed to release the entity when the simulated episode duration has elapsed. After its release, the entity moves down a model path, through various "assign" and "record" modules in order to update and record the new status of its attributes until it reaches another "queue", in this instance one that represents one of the recovered health states.

The grey boxes containing zeros in Figure 2 are windows presenting the value of important output variables. These windows are useful since they allow the status of the simulated population to be monitored while a simulation is in progress. However, they play no role in the actual simulation.

Figure 2. Overview of the Simulation Model.



Section 1: Initial Assignments

Figure 3 identifies the part of the model in which initial assignments are made. In the model, entities are generated (in a "generate" module) and are then assigned attributes (by an "assign" module) at baseline. Figure 4 presents an expanded depiction of this part of the model.

Figure 3. Entity Generation and Initial Assignments

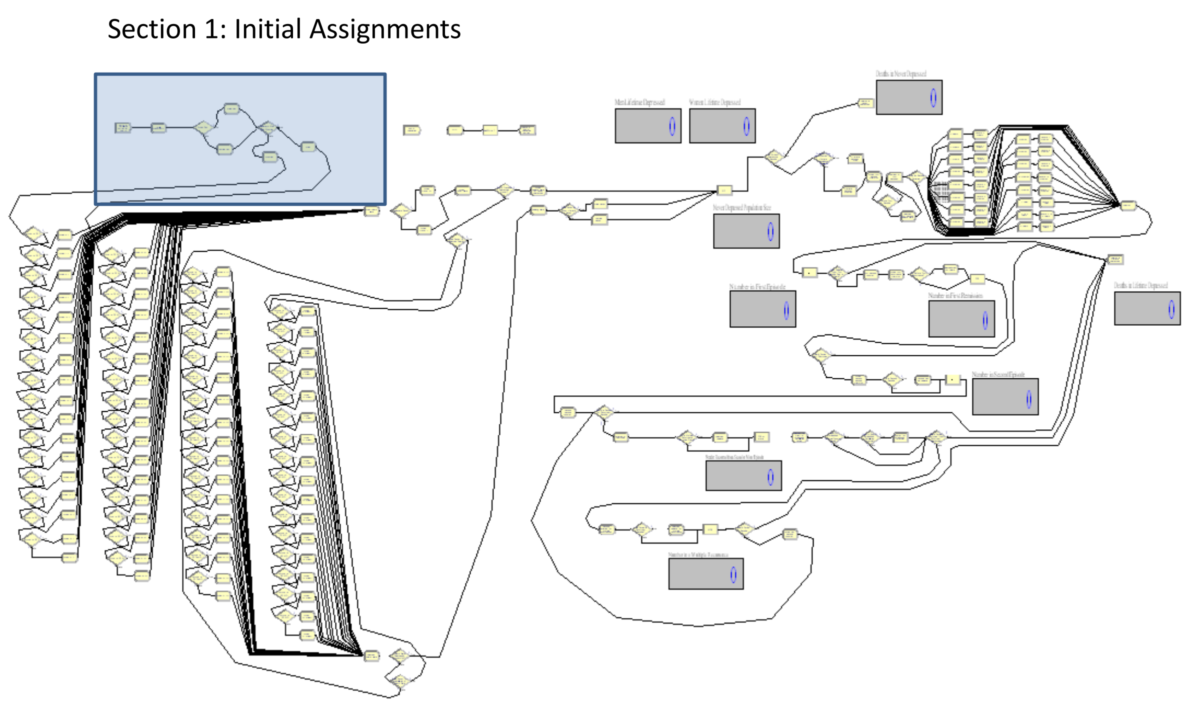
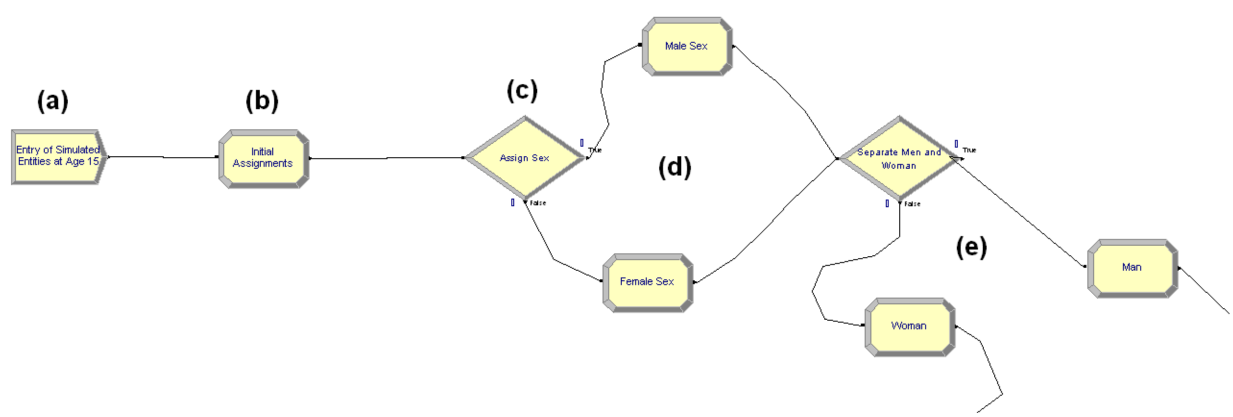


Figure 4. Zoomed Screen Capture of Section 1 in the Model.



(a) A “generate” module that randomly generates new entities at a predetermined rate simulated using an exponential parameter reflecting the average time between entity entries. This is the reciprocal of the rate of entry of new attributes.

(b) An “assign” module that assigns entity attributes as each entity enters the simulation. The assignments made in this module are presented in Table 1. As noted above, each entity is assigned an age of 15 at the time of entry.
(c) A “decide” module that randomly assigns an attribute for sex to each entity – the probability of being male is set at 50.5%.
(d) The attribute depicting sex is assigned to the entities in these “assign” modules.
(e) The male and female entities are divided in preparation for entry into Section 2 of the model, in which a value (attribute) for their simulated length of life and date of death is assigned (see Section 2, below).

Table 1. Assignment of Attributes in the Initial “Assign” Module (b)

Name	**Attribute or Variable**	**Role in the Model**
Entry Date*	Attribute	Identifies time of entry into the simulated population.
Age at Death Category	Attribute	The attribute is defined at this point, but is not given a specific value. Vital Statistics data are later used to assign a death date to each entity (see Section 2, below) in 5-year age ranges.
Death into Interval	Attribute	The attribute is defined at this point, but is not given a specific value. Later a variable (generated from a uniform random distribution) is used to assign an exact date of death within the 5-year range where the death occurs.
Death Date	Attribute	The attribute is defined at this point, but is not given a specific value. Later, a death date is simulated by adding the simulated duration of life to the date of entry into the model (see Section 2)

* time is depicted in the model in days, with the value of the SIMAN system variable “TNOW” being assigned to each entity as an attribute at the time of entry into the model.

Section 2: Assignment of Life Expectancy

Figure 5 identifies the part of the model concerned with assignment of life expectancy to each entity. Simulation of life expectancy for each model entity is essential since mortality drains the prevalence pool and contributes to the establishment of a steady state population size. Following the approach taken by Kruijshaar et al. (3), after assignment of a death date, an onset date for a first depressive episode is simulated (using a Weibull distribution, see Section 3). If the onset date falls after the entity's simulated death date, then no episode occurs during that entity's lifespan. For purposes of simulating life expectancy, vital statistics data from Canada's national statistical agency (Statistics Canada) were used. Up to date mortality tables in 5-year age groups were obtained and the occurrence of death was simulated based on these probabilities. When a death occurs, a random value from a uniform probability distribution is generated and used to assign a specific death date within the five year interval, see "Death into Interval" in Table 1.

Figure 5. The Part of the Model Concerned with Simulation of Life Expectancy.

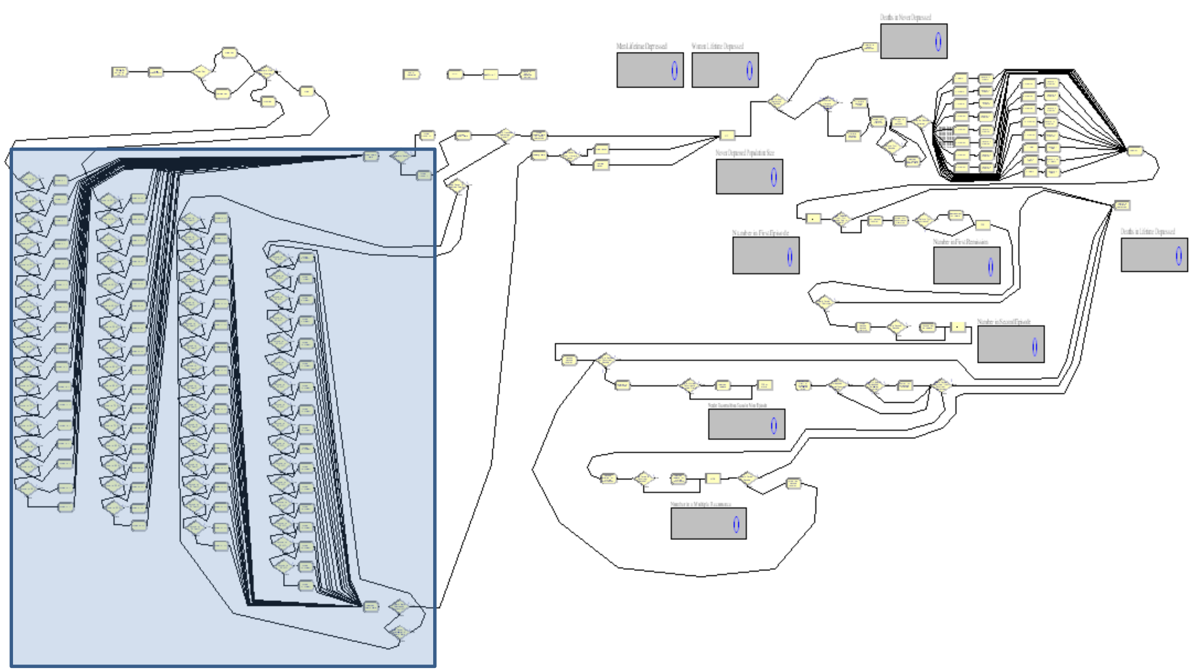
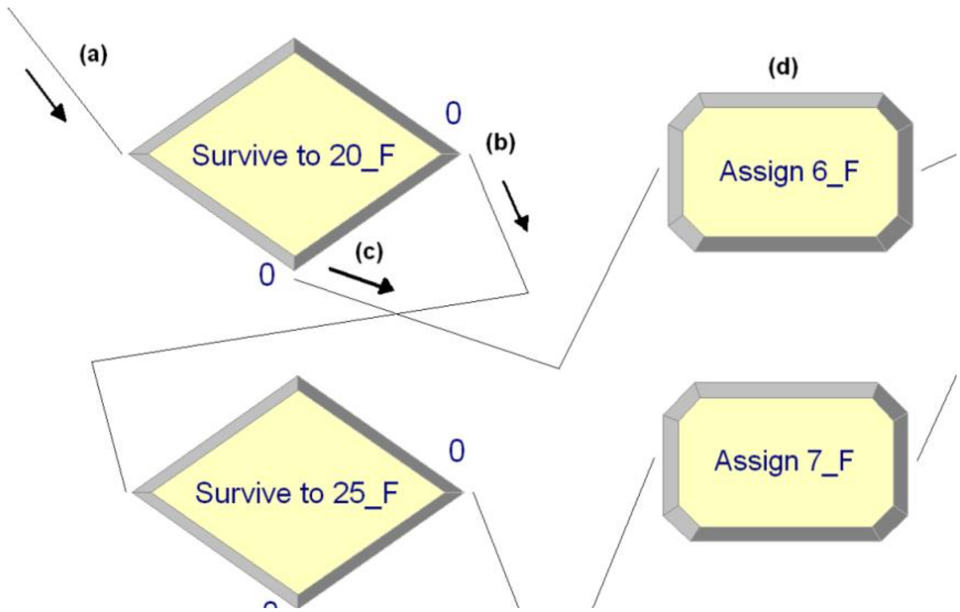


Figure 6 depicts a representative part of the process for assigning life expectancy. Entities enter the model at age 15, depicted as (a) in Figure 6. Initially, their probability of death during the 15-20 year age range is simulated using a "decide" module, with that probability being determined from vital statistics data, as described above. Those who survive (b) move on the next "decide" module, which repeats the process for the next five year age category. Those who do not survive the five year interval are moved on to an “assign” module which assigns a date of death within this five year interval (d) using a randomly generated death into interval variable from a uniform distribution (see Table 1).

Figure 6. Enlarged Segment of Section 2 in the Simulation Model*



* The “F” refers to female, the model contains a similar but separate (see Figure 1) mechanism for assignment of death dates in men, the latter using vital statistics data for Canadian men.

(a) A female entity that has survived from entry (age 15 to age 20 enters this part of the simulation following this path.

(b) This entity survives this age range and moves on to the next 5 year age interval.

(c) This entity dies during this age range and moves along this path to “assign” modules that assign a death date.

(d) A specific death date within the five year age interval is assigned using a uniform random distribution which is then converted to the proportion of the interval survived and subsequently to a specific date of death.

When a death date has been assigned to an entity, it moves on to Section 3 of the model (see below), where an incidence date is assigned. Because major depression may be associated with an increased risk of mortality, e.g. see (4), the model includes a variable depicting relative risk for mortality. When a major depressive episode has been assigned to an entity in Section 3, they are returned to Section 2 and the life expectancy is recalculated with each age specific mortality probability being multiplied by this relative risk. For example, if the relative risk of mortality were set to 1.4, then each age specific mortality risk in Section 2 is increased by 40% for entities with lifetime episodes. For the simulations presented in this paper, the relative risk for mortality was set to 1.0, except in a sensitivity analysis carried out during calibration of the model, as described below.

Section 3: Simulation of the Incidence

Figure 7 displays the part of the model involved in the simulation of incidence. As noted above, the simulation was based on whether the simulated time to first episode was less than the death date previously assigned. The interval to first episode was simulated using two Weibull distributions, one for male and one for female entities. The Weibull distribution describes a rate that changes over time. Each distribution was specified using separate “scale” and “shape” parameter, thereby specifying a two-parameter Weibull distribution. If a single rate was anticipated, then an exponential distribution could have been used for this part of the simulation, but the use of the Weibull distribution allows the incidence rate to decline with the simulated age of the entity, as is known to occur in major depression (5). SIMAN uses an inverse form of the cumulative Weibull distribution in order to simulate this length of time starting with a simulated value from a uniform distribution.

Figure 7. The Part of the Model Concerned with Simulation of First Incidence

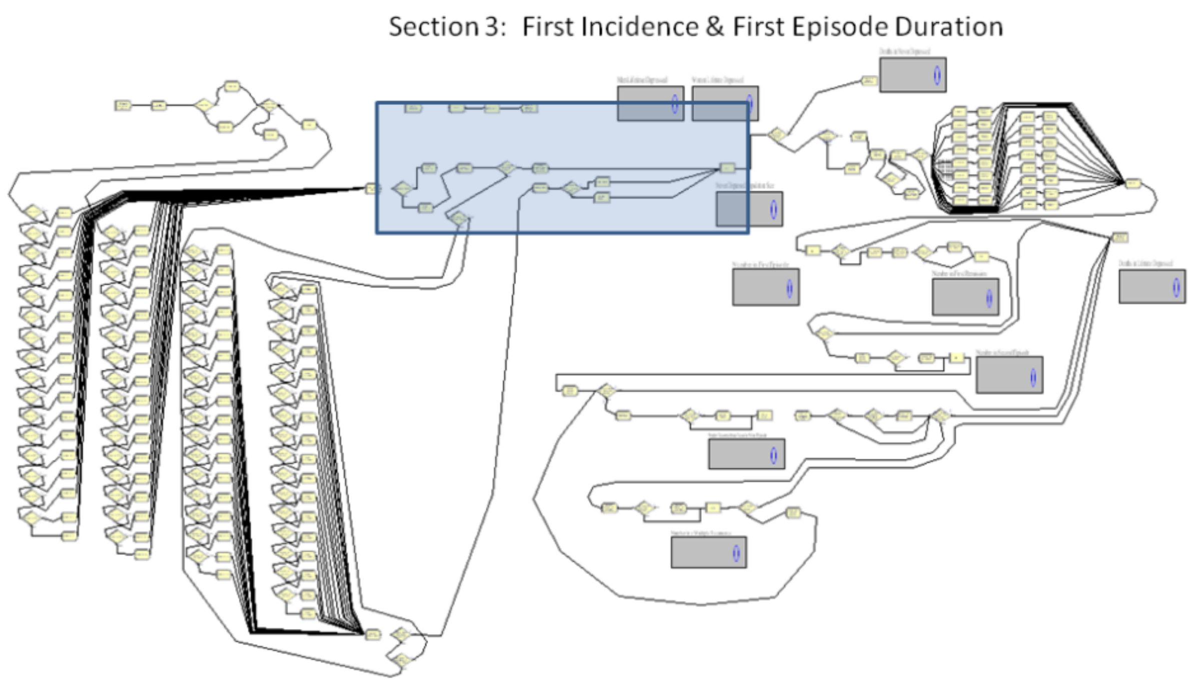
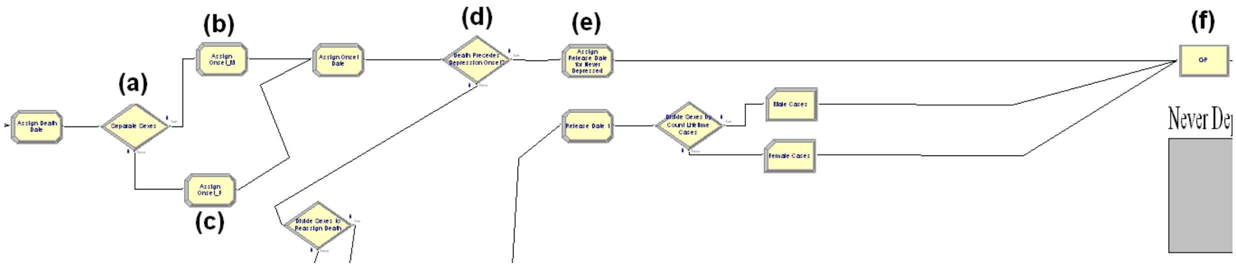


Figure 8. Expanded View of the Modules for Assignment of Initial Episodes*



(a) A “decide” module dividing the sexes prior to simulation of a first episode onset date.
(b) An “assign” module to assign a first episode onset date, in men.
(c) Assignment of a first onset date in women.
(d) This “decide” module identifies those with episodes (lifetime prevalence positive) and sends them back to Section 2 in order to reassign a death date (see above).
(e) Assigns a “release” date from the “general” population, ie. those members of the simulated population who have not yet experienced an episode – the release date determines the date at which those respondents who develop an episode during their lifetime leave the queue representing this segment of the population. For those with lifetime episodes (first episode onset date precedes the death date), the release date is the onset date for their first episode, whereas for those without lifetime episodes the release date is the simulated date of their death.
(f) A “queue” storing simulated entities in this “general population” (GP) compartment, see Figure 1. The queue is scanned by SIMAN for the following condition: simulation time (TNOW) > release date, see (e) above. When this condition is true, the entity in question is released from the queue.

Section 4. First Episodes and Associated Recovery

When an entity has an initial episode, it is released from the general population (GP) queue, as described above, and moves into a first episode (FE) queue, see Figure 1. The entity is processed initially by the assignment of a simulated episode duration for the first episode. This duration is again simulated using a Weibull distribution also characterized by a shape and scale parameter. This represents the tendency of the recovery rate in major depression to decline with increasing episode duration (6). The same Weibull distribution was used to simulate the duration of episodes in male and female entities. Each entity is assigned a release date from the first episode queue based on the entity’s time of entry and the duration of the episode. The model has a module to check whether the simulated death date occurs before the resolution of the episode. If so, the release date is reassigned so that such entities are released from the queue at their time of death rather than at their time of recovery. An additional module is present on the model path after the queue in order to determine whether the release was due to recovery or death. The entity then proceeds either to the first recovery state (see Figure 1), or to a dispose module, depending on their reason for release from the queue. Figure 9, below identifies the part of the model concerned with these transitions. A more detailed view of this part of the model is presented in Figure 10, below.

Figure 9. The Part of the Model Concerned with First Episodes and Associated Recovery

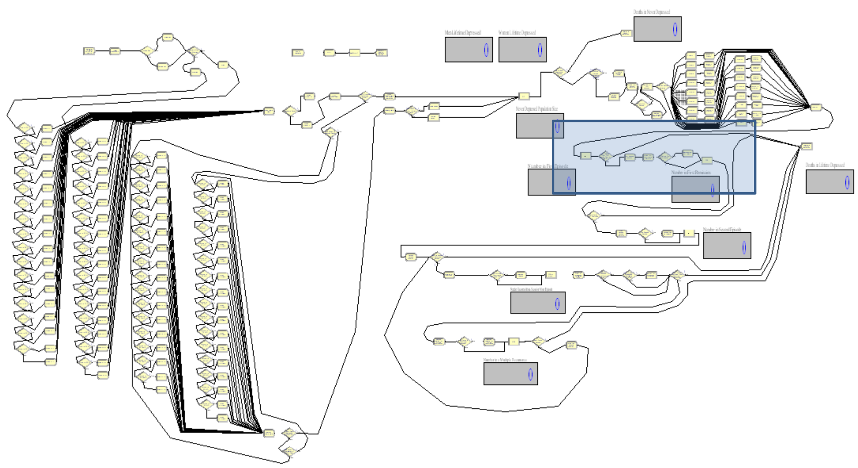
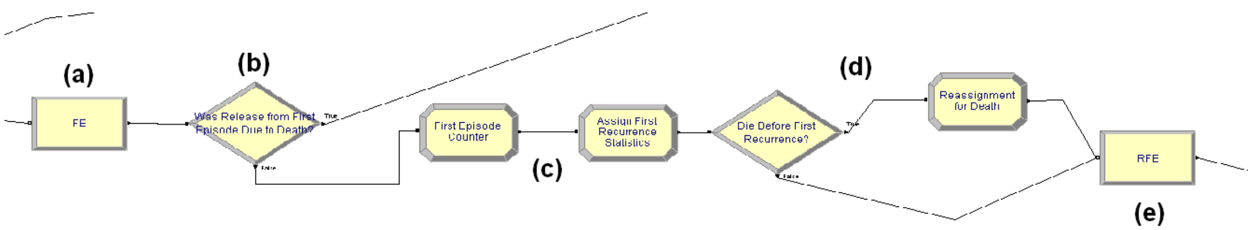


Figure 10. Expanded View of the Modules for First Episodes and Associated Recovery



(a) The queue containing entities in their first episode.

(b) A "decide" module to detect those who were released from the first episode due to death rather than recovery. After entity values are recorded, these entities move on to the "dispose" module representing their death.

(c) "Assign" modules for tracking statistics and assignment of an attribute representing the duration of the remission period following the first episode.

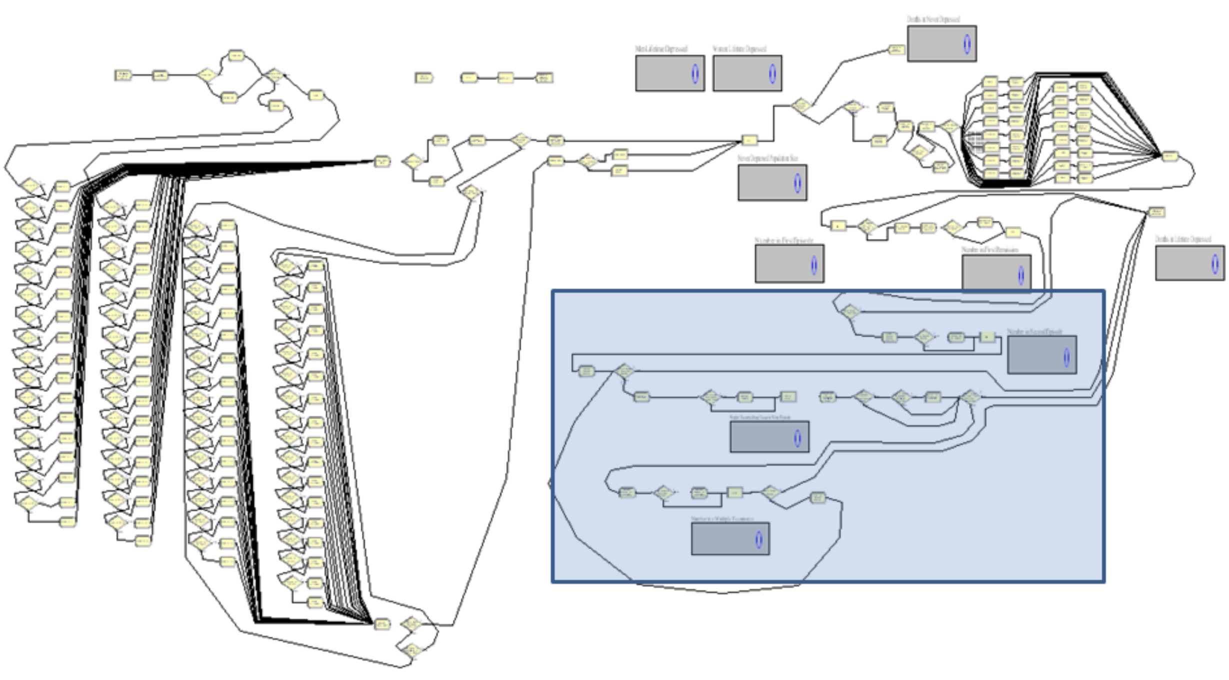
(d) A "decide" module to detect those entities who recover from a first episode but who die before having a second episode. The release date from the recovered from first episode (RFE, see Figure 1) queue is then re-assigned to the simulated death date.

(e) This queue holds the entities who have recovered from the first episode, but whose release date (either by recovery or death) from that queue has not been reached.

Section 5. Second and Multiple Episodes and Associated Recovery

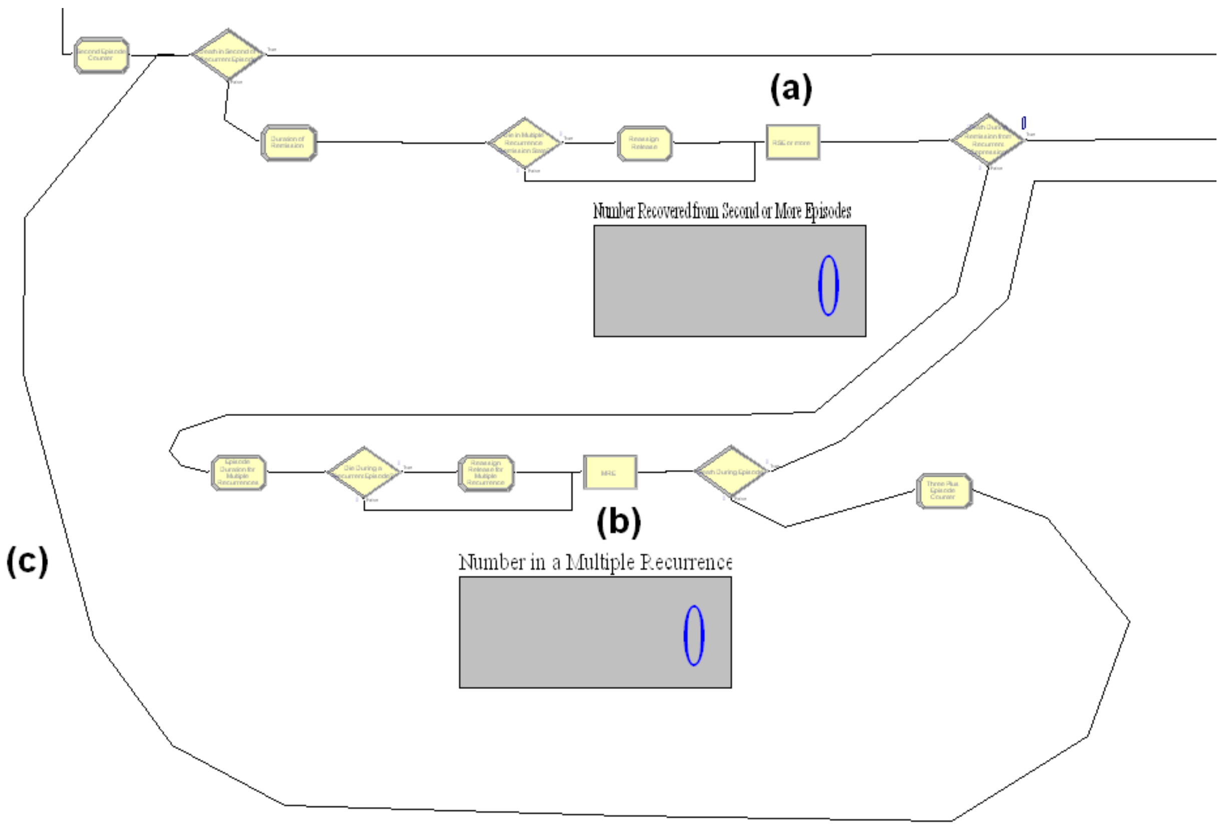
The general approach to simulation of second episodes was the same as that of first episodes. Entities are released from the queue containing those who have recovered from an initial episode using the same general approach: assignment of a release date based on simulation of the duration of time spent in the state. The release date corresponds either to the emergence of new episode, or death, whichever occurs first in the simulation, see above. The duration of both first and second recurrence, however, were simulated using an exponential distribution rather than the Weibull distribution. Two different exponential distributions were used, one for first recurrence (ie. recurrence after a single episode) and one for multiple recurrences (ie. recurrence after two or more episodes). This may be a simplification (the rate of recurrence may decline with time spent in remission), but reduces the number of parameters needed for calibration of the model. Figure 11 identifies the part of the model concerned with simulation of second and multiple episodes.

Figure 11. The Part of the Model Concerned with Second and Multiple Episodes, and Associated Recovery



Each of the relevant health states: second episodes (SE), recovered from second (or more) episodes (RSE or more) and multiple recurrent episodes (MRE) are represented as queues in the model, with release dates specifying the simulated date of movement into the next state because of recurrence, death or recovery. The queue for those who have recovered from a second (or more) episode includes those entities who have recovered from multiple episodes as well as from second episodes because the final step in the model is a loop from the multiple episode state, returning to the queue for recovery from a second or more (in this case more) episode(s), as depicted in Figure 11.

Figure 11. Expanded View of the Part of the Model Depicting Second and Multiple Episode and Associated Recovery.



(a) Queue for entities who have recovered from a second (or more) prior episode(s).
(b) Queue for entities experiencing a third, or more, recurrence.
(c) Path followed by entities following recovery from multiple (3 or more) episodes, back to queue (a) after confirmation that they did not die in the second episode, simulation of the duration of their remission and assignment of a release date from the queue based on the duration of their remission or their death.

Simulation of Epidemiologic Parameters

For purposes of calibration, the simulation model must produce output that can be compared to estimable parameters from the CCHS 1.2 survey. Table 2 (in Part 4, below) lists the four variables upon which calibration was based. The first is current prevalence. In the CCHS 1.2, this is assessed by a question about the most recent episode. This was asked to all respondents with a history of recurrent major depressive disorder. They were classified as having current major depression if they reported being in an episode during the preceding 30 days. In the model based simulation, current prevalence was calculated using variables representing the number of entities in each queue. The number of entities in the first episode, second episode and multiple recurrent episode queues was divided by the total number of entities in the simulation at the same point in time. Another parameter used for calibration was lifetime prevalence. Since each entity was assigned an attribute for sex, a variable counting the number of entities in each queue having the female or male attribute was used in the calculation of lifetime prevalence for men and women. The sum of those in an episode (same as current prevalence above) and those in the two recovery states: recovered from a first episode (RFE) or recovered from a second, or more (RSE or more) episode were also included in the numerator. The denominator of the lifetime prevalence calculation was the total number of female or male entities in the simulation at a point in time. Additional items from the CIDI sought information about the duration of first episodes and the total number of episodes lasting two weeks for more in a respondent's lifetime. Unfortunately, the CCHS 1.2 did not collect data about the duration of episodes in general. For this reason, the model had to assume that recurrences had the same prognosis as first episodes. The duration of first episodes was a simulated parameter (see Section 4, above), based on the Weibull distribution. The number of episodes was depicted by an attribute assigned an initial value of 0 and which then counted up by 1 with the occurrence of each new episode.

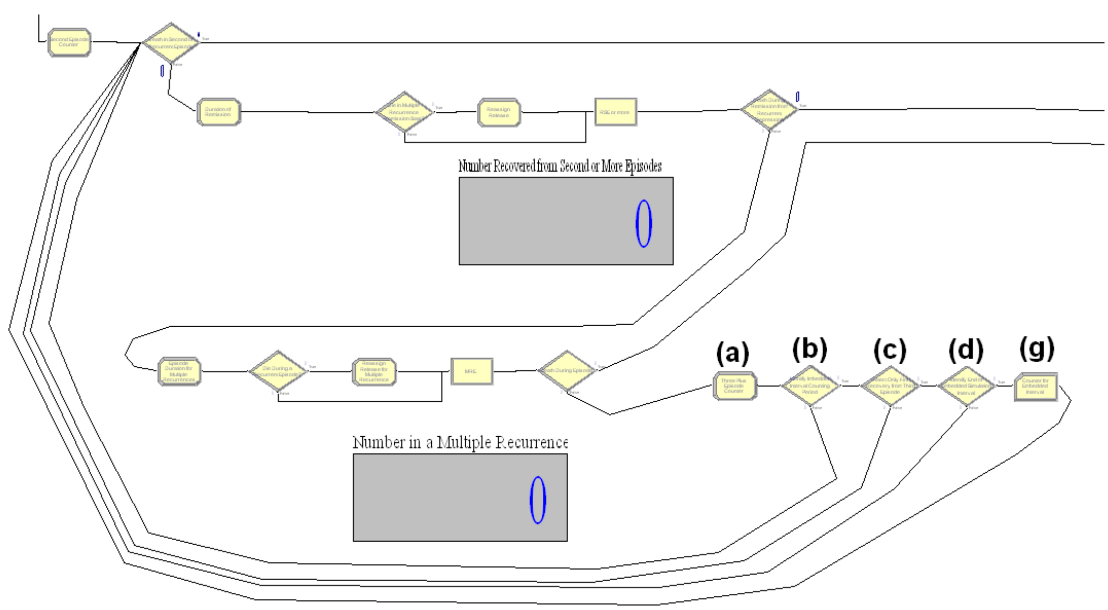
Part 2. Adaptation of the Simulation Model for Calibration with the CCHS 1.2 Dataset

The general approach to simulation described above includes a depiction of the recurrence rate after recovery from an initial episode and for that after recovery from two or more episodes. Calibration of the exponential distributions used to simulate these rates would ideally use a longitudinal data source. However, the CCHS 1.2 is a cross-sectional study. In order to use the CCHS 1.2 estimates to calibrate the model it was necessary to depict the model output in a way that resembled a cross-sectional dataset rather than the dynamic simulation described above. To accomplish this, a "sampling date" was added to the model at steady state (150,000 days into the simulation, or approximately 400 years). All of the queues in the model were programmed to release their entities on that date and additional modules were added in order to record various attributes (including the number of episodes experienced by those entities as of that particular date) upon release of the entities from whatever queue they occupied at the time of the sampling date. This created a dataset of the type that would derive from a cross-sectional survey, allowing calculation of the four key variables: current prevalence, sex-specific lifetime prevalence, episode duration and number of episodes as a "snapshot" at a particular date analogous to a cross-sectional study.

Part 3. Adaptation of the Model to Assess Incidence of Three or More Episodes

Additional Modules were added to the part of the model depicted in Figure 11 in order that the simulation output would include a measure of the incidence of three or more episodes, a central parameter in view of the goals of the project in relation to Mindfulness-based Cognitive Behavioural Therapy. These additions are depicted in Figure 12. At steady state, a 10 year period was identified in order to calculate a 10 year incidence proportion for recovery from a third episode. First a "decide" module was used to identify entities recovering from a third (or more) episode during this 10 year period, in distinction to the basic structure of the model, which does not single out this group. A "record" module was then used to count the specific entities whose number of prior episodes was three at the time of this recovery event. The denominator for the calculated incidence proportion was the sum of the number of entities in all of the model's queues at the start of this simulated 10 year interval.

Figure 12. Additions to the Part of the Model Depicted in Figure 11 to Count the 10-year Incidence Proportion of Recovery from a Third Episode.



(a) "Assign" module adding a new past episode to those entities recovering from a recurrent episode. Each entity possessed this episode counting attribute to which a value of one is added during each passage through this module.

(b) “Decide” module to remove entities prior to the selected 10 year measurement interval since these are not used in simulation of the targeted rate.
(c) “Decide” module to include only those recovering from their third prior episode, enabling a count of the number of recoveries from three episodes.
(d) “Decide” module removing entities who recover after the end of the selected 10 year measurement interval from the calculation of the targeted rate.

Part 4. Model Calibration

Table 2 summarizes the approach to calibration. Each of the simulated parameters can be considered to be primarily determined by one variable. For example, lifetime prevalence is primarily determined by incidence (and mortality), since episode duration or the number of recurrent episodes do not influence lifetime prevalence. Similarly, episode duration is determined primarily by recovery rates. The number of episodes is determined primarily by recurrence rates. However, the number of episodes is likely also to be determined by the age of onset of first episodes, since this affects the time at risk for recurrence. Point prevalence is jointly determined by all of the other factors. The relative risk for mortality, in theory, could also impact each of the simulated parameters, but this was set to 1.0 in the simulations presented here, except for a sensitivity analysis specifically addressing the issue of mortality.

The model was calibrated by running a large (n=10,000) number of simulations across a range of plausible values for each of the model variables, with the objective of minimizing the squared differences, or sum of squared differences, between simulated and observed epidemiologic parameters, see Table 2.

In each series of simulations, the targeted model parameter was allowed to vary across a specified range. The output of each simulation was recorded as the square of the difference between the simulated and estimated output parameter. For example, in order to identify the Weibull scale and shape parameters leading to the best description of CCHS 1.2 episode duration, the simulated proportions reporting episodes lasting each of 16 reported durations (ranging from 2 weeks to 5 years or longer) from the CCHS 1.2 were subtracted from the proportions reporting each of these 16 durations in the CCHS 1.2 dataset. Each of these differences were squared so that negative differences would not offset positive ones and the sum of squares was calculated for each simulation. Automated software packaged with the Arena® simulation program, called OptQuest®, was used to automate the calibration process. OptQuest begins with a range of values and, as the simulation proceeds, seeks to focus on an increasingly narrow range of values to accomplish a calibration objective. In this case, the calibration objective was to minimize the squared differences or sum of squared differences resulting from the simulation runs.

Table 2. Approach to Model Calibration

	Epidemiologic Parameters Primarily Determining this Variable*	Model-based Parameters	Calibration Standard
Episode Duration**	Recovery Rates	Weibull Scale and Shape Parameters	(Minimize) sum of squares of simulated minus reported proportions with 16 episode durations (first* episodes).
Lifetime Prevalence	Incidence (sex specific)	Weibull Scale and Shape Parameters for sex-specific incidence	(Minimize) the squared difference: simulated minus observed lifetime prevalence in men and women.
30 Day (current) Prevalence	Incidence, Recovery and Recurrence	Parameters reflecting Incidence, Recovery and Recurrence	(Minimize) the squared difference: simulated point prevalence minus observed 30-day prevalence from the CCHS 1.2
Recurrence	Recurrence Rates for Recurrence after a single (a) or multiple (b) prior episodes.	Two exponential parameters (a) and (b)	(Minimize) the sum of squares of simulated proportion with each number of episodes minus the reported proportions with that number of episodes in the CCHS 1.2

* The CCHS included items asking about the duration (in weeks, months and years) of first episodes. The values were converted to duration in days for the calibration and the same Weibull parameters were used to simulate recurrent episodes.

Once a range of seemingly plausible values for each parameter were identified, additional simulations were carried out in which OptQuest was allowed to vary each of the model parameters listed in Table 2 and to identify simulations that would minimize the sum of squared differences across all of the various outputs. The ranges of values employed in these final calibration runs of n=10,000 simulations are presented in Table 3. Note that OptQuest also allows the specification of a “suggested value” and the suggested values used in the simulations are also presented in Table 3.

Table 3. Values Selected from Preliminary Simulations for Inclusion in the Final Model Calibration.

	Model Based Parameters	Pre-calibration Values		
		Low Bound	Suggested Value	High Bound
Episode Duration*	Weibull Scale	210	230	240
	Weibull Shape	0.41	0.46	0.49
Incidence	Weibull Scale (men)	610,000	646,000	680,000
	Weibull Shape (men)	0.6	0.67	0.73
	Weibull Scale (women)	1,000,000	1,161,707	1,300,000
	Weibull Shape (women)	0.39	0.44	0.49
First Recurrence	Exponential Parameter**	10,000	11,000	12,000
Multiple Recurrence	Exponential Parameter**	3300	3900	4500

* as 14 days is the minimal duration by definition, the duration of episodes was simulated as 14 days plus a simulated value from the Weibull distribution.

** these parameters reflect the average interval, in days, between episodes. They are the reciprocal of the rates of recurrence with days^{-1} units.

Part 5. Results of the Model Calibration

Table 4, below, lists the actual values selected by OptQuest for the various model parameters in the final set of simulations. In the case of each parameter, a relationship showing evidence of a parameter value that minimized the squared differences or sum of squared differences was evident. This is depicted in a series of scatter plots, each containing a polynomial regression line in order to highlight the observed pattern. Each scatterplot represents n=1000 simulations in which each parameter is altered while holding the others constant. These scatterplots are depicted in Figures 13 through 20. The results that follow are from simulation runs using an exponential parameter of 5 for entry, meaning that the average simulated time interval between entry of new entities was 5 days. Given the applicable mortality data, this resulted in a simulated steady state population of approximately 6000. As such, the estimates are vulnerable to some degree of random error.

Table 4. Results of the Model Calibration

	Model Based Parameters	Calibrated Values
Episode Duration*	Weibull Scale	220.4
	Weibull Shape	0.478
Incidence	Weibull Scale (men)	610000
	Weibull Shape (men)	.609
	Weibull Scale (women)	1074360
	Weibull Shape (women)	0.430
First Recurrence	Exponential Parameter**	10961
Multiple Recurrence	Exponential Parameter**	3862

* as 14 days is the minimal duration by definition, the duration of episodes was simulated as 14 days plus a simulated value from the Weibull distribution.
** these parameters reflect the average interval, in days, between episodes. They are the reciprocal of the rates of recurrence with days^{-1} units.

Figure 13. Sum of Squared Differences Between Observed and Simulated Episode Durations: Weibull Scale Parameter Values Ranging from 50 to 500 (N=1000 Simulations)

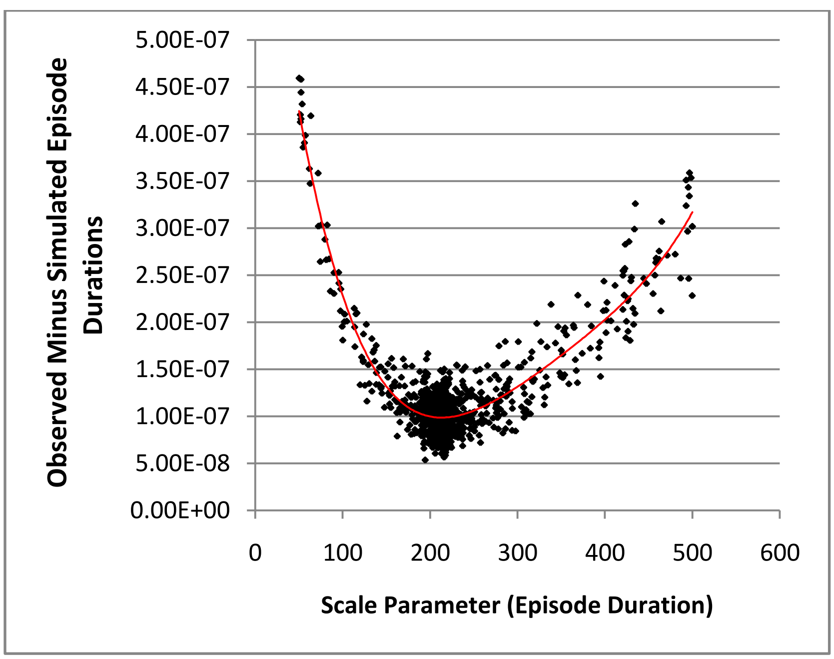


Figure 14. Sum of Squared Differences Between Observed and Simulated Episode Durations: Weibull Shape Parameter Values Ranging from 0.1 to 0.9 (N=1000 Simulations)

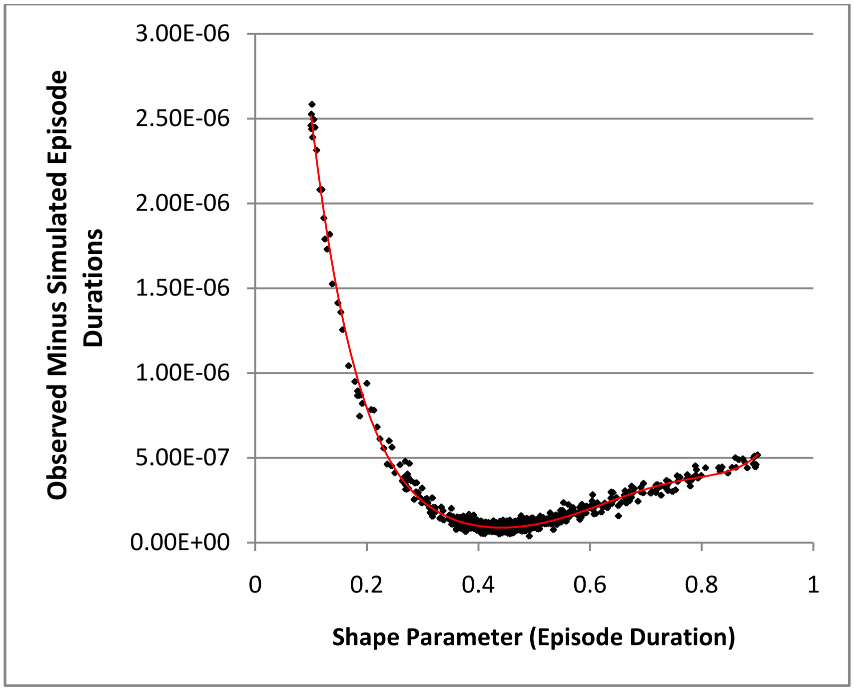


Figure 15. Squared Difference Between Observed and Simulated Lifetime Prevalence (Male): Weibull Scale Parameter Values Ranging from 250,000 to 2,000,000 (N=1000 Simulations)

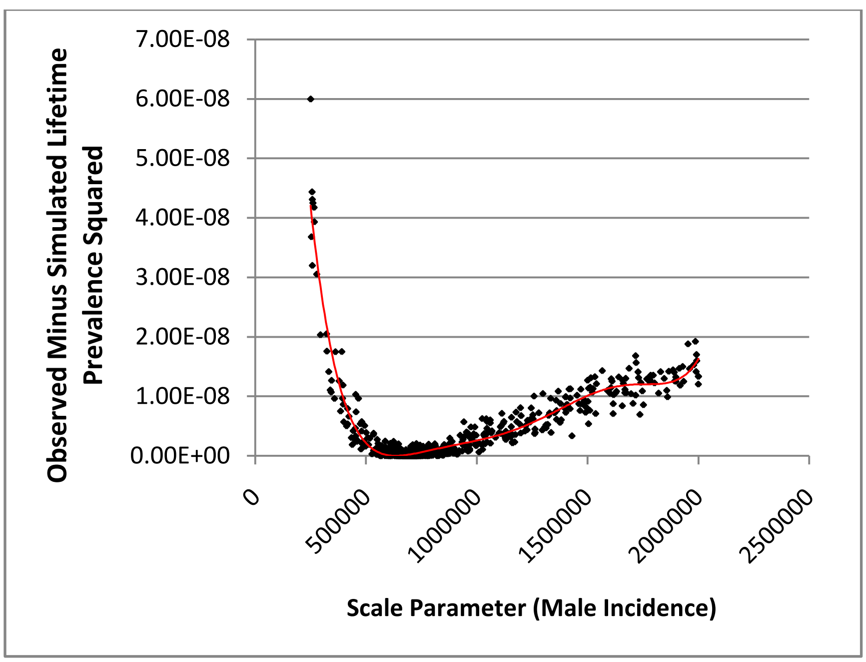


Figure 16. Squared Difference Between Observed and Simulated Lifetime Prevalence (Male): Weibull Shape Parameter Values Ranging from 0.4 to 1.0 (N=1000 Simulations)

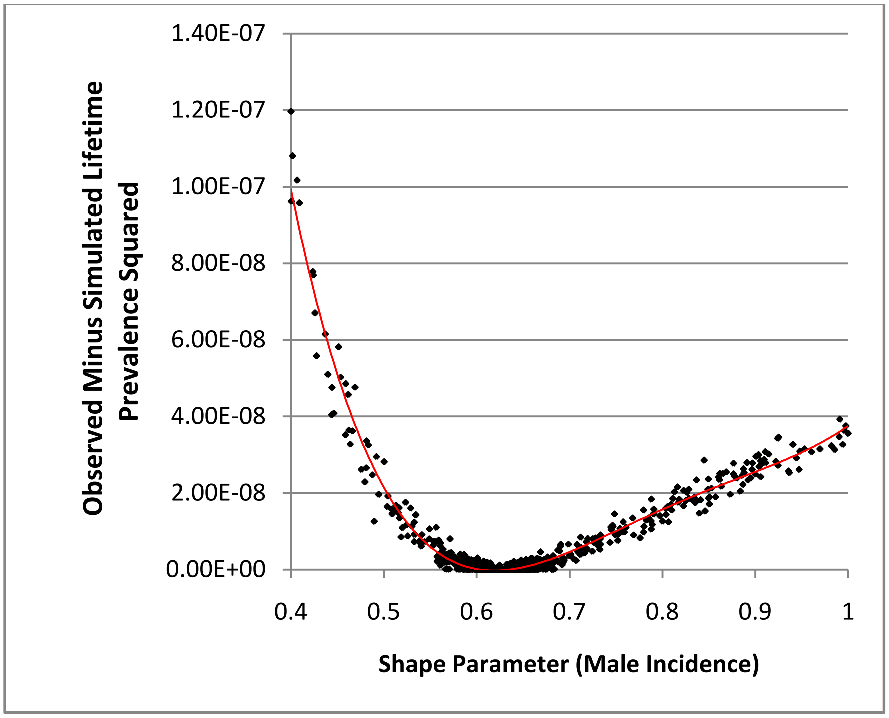


Figure 17. Squared Difference Between Observed and Simulated Lifetime Prevalence (Female): Weibull Scale Parameter Values Ranging from 500,000 to 2,000,000 (N=1000 Simulations)

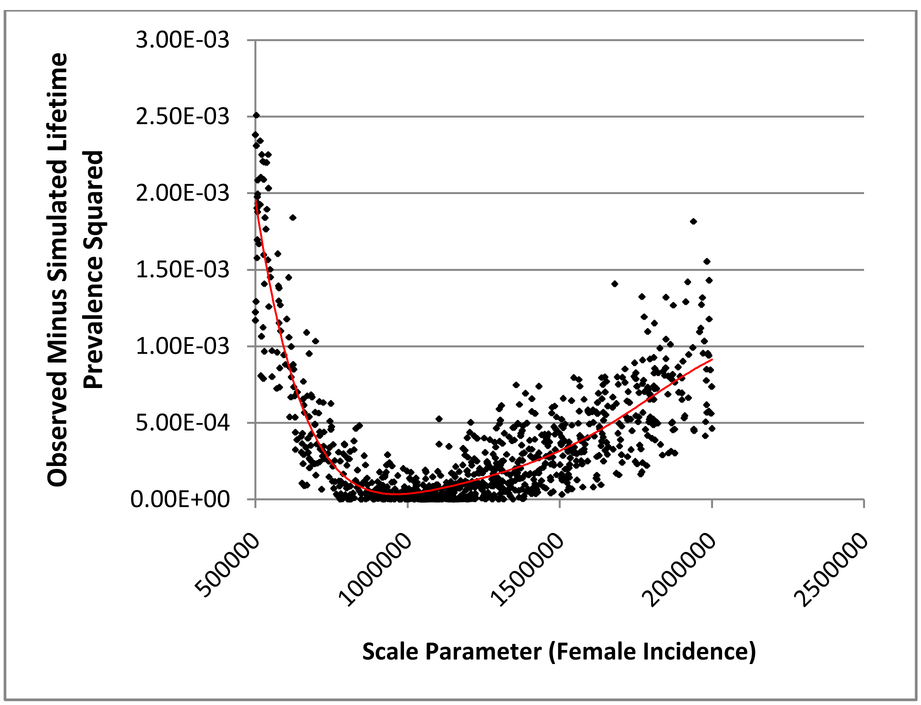


Figure 18. Squared Difference Between Observed and Simulated Lifetime Prevalence (Female): Weibull Shape Parameter Values Ranging from 0.1 to 0.9 (N=1000 Simulations)

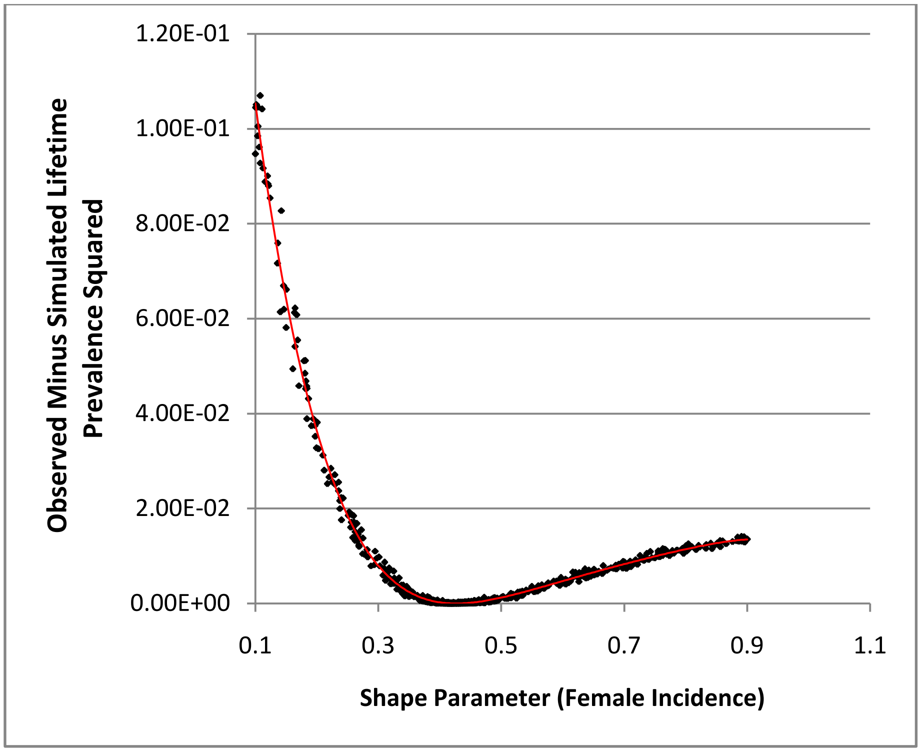


Figure 19. Sum of Squared Differences Between Observed and Simulated Number of Episodes: Exponential Parameter Values Ranging from 5000 to 25,000 (N=1000 Simulations)

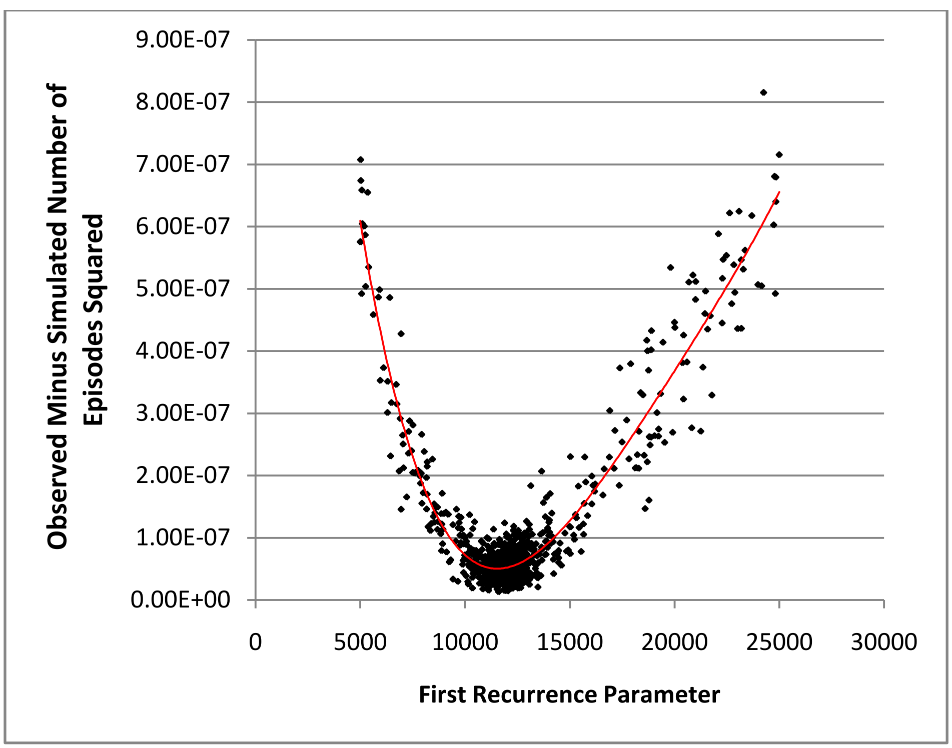


Figure 20. Sum of Squared Differences Between Observed and Simulated Number of Episodes: Multiple Recurrence (Exponential) Parameter Values Ranging from 1000 to 10,000   (N=1000 Simulations)

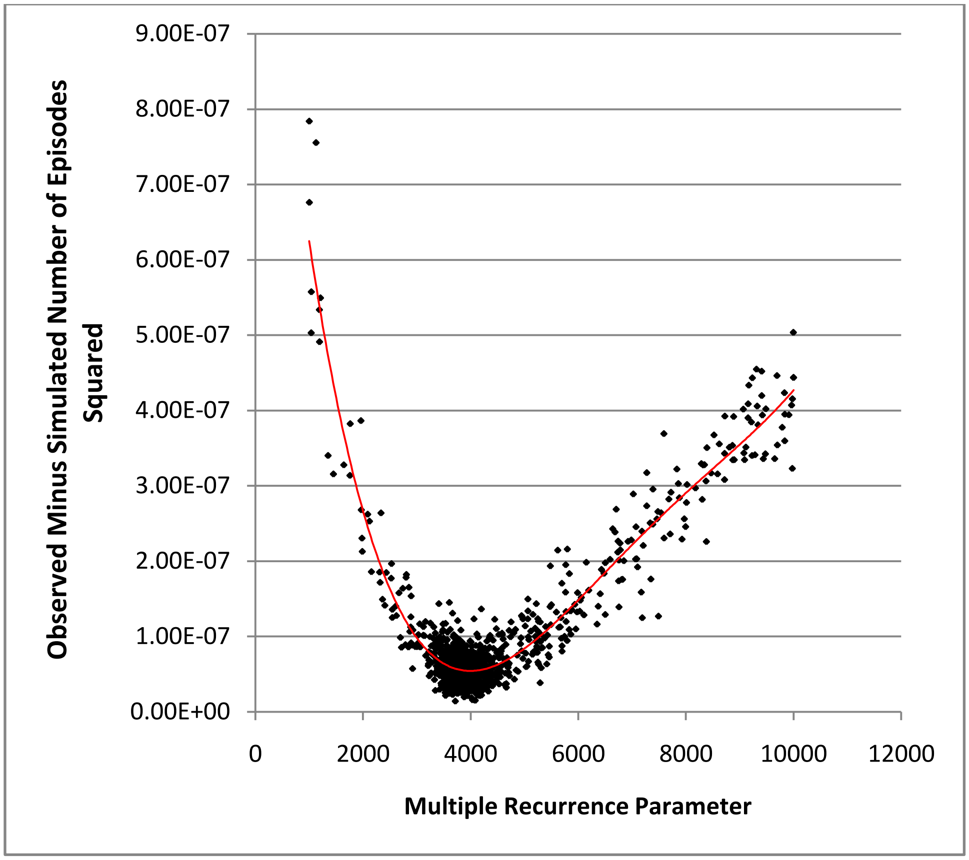


Figure 21 presents the average results of 1000 simulation runs for current prevalence and sex-specific lifetime prevalence. Figure 22 presents actual and simulated (average of n = 1000 runs) data for number of past episodes. Histograms representing the same three output parameters after n=1000 simulation runs are presented in Figures 23-25. Figure 26 presents the reported cumulative proportion reporting recovery, by week, from the CCHS 1.2 along with the average values from the 1000 simulations runs. The main point of deviation occurs around the 6-month time point, presumably because six month episode durations are preferentially reported as compared, for example, to five or seven month episode durations. Such “rounding” is not surprising since the first episode duration question in many cases would have referred to an episode occurring many years earlier. This effect is more evident when the results are presented in a non-cumulative distribution. Figure 27 presents the proportion recovering by week (in this case the cumulative Weibull function from the parameters in Table 4) juxtaposed against the self-reported episode durations. There is a predominance of episodes having a reported duration of six months. The outlier at six months is also evident on a quantile-quantile plot (Figure 28).

Figure 21. Simulation Model Output Using the Calibrated Parameters (see Table 4), Juxtaposed with Estimates from the CCHS 1.2

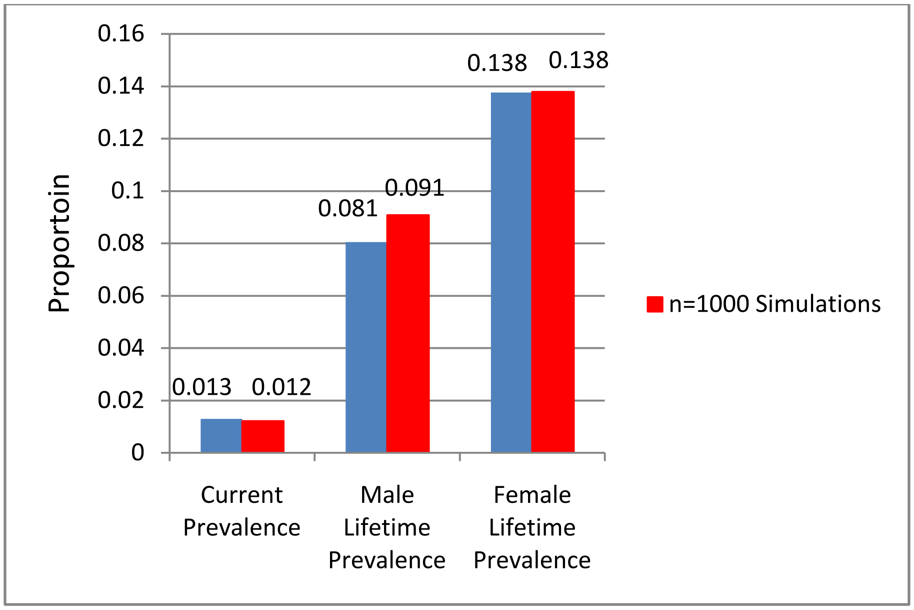


Figure 22. Actual (CCHS 1.2 Estimates) and Simulated Number of Past Episodes

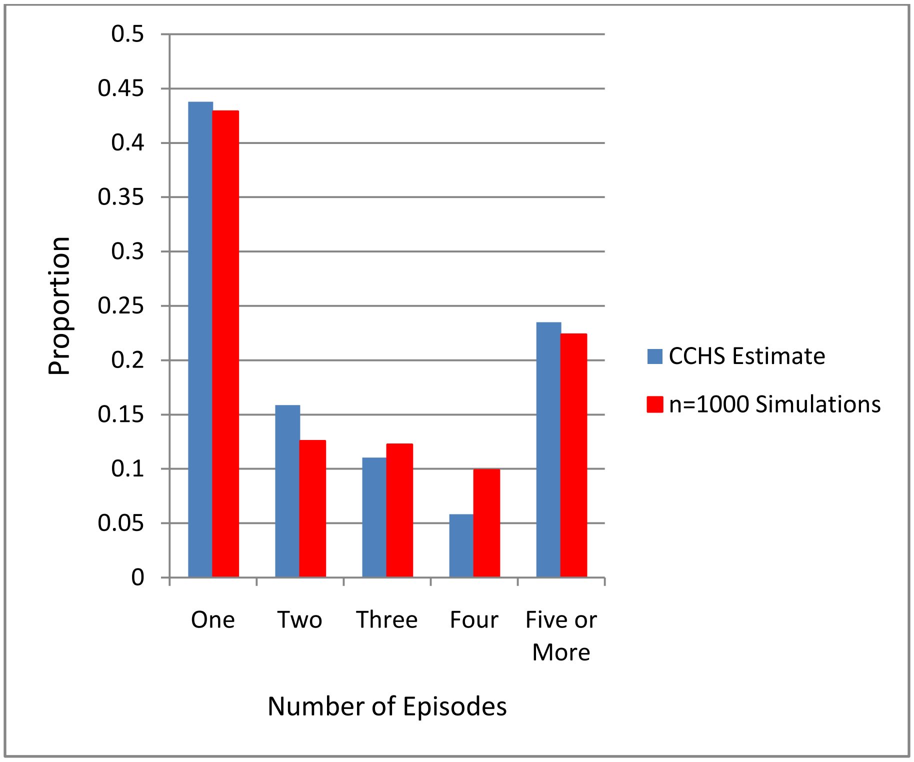


Figure 23. N=100 Simulated Estimates of Current Prevalence

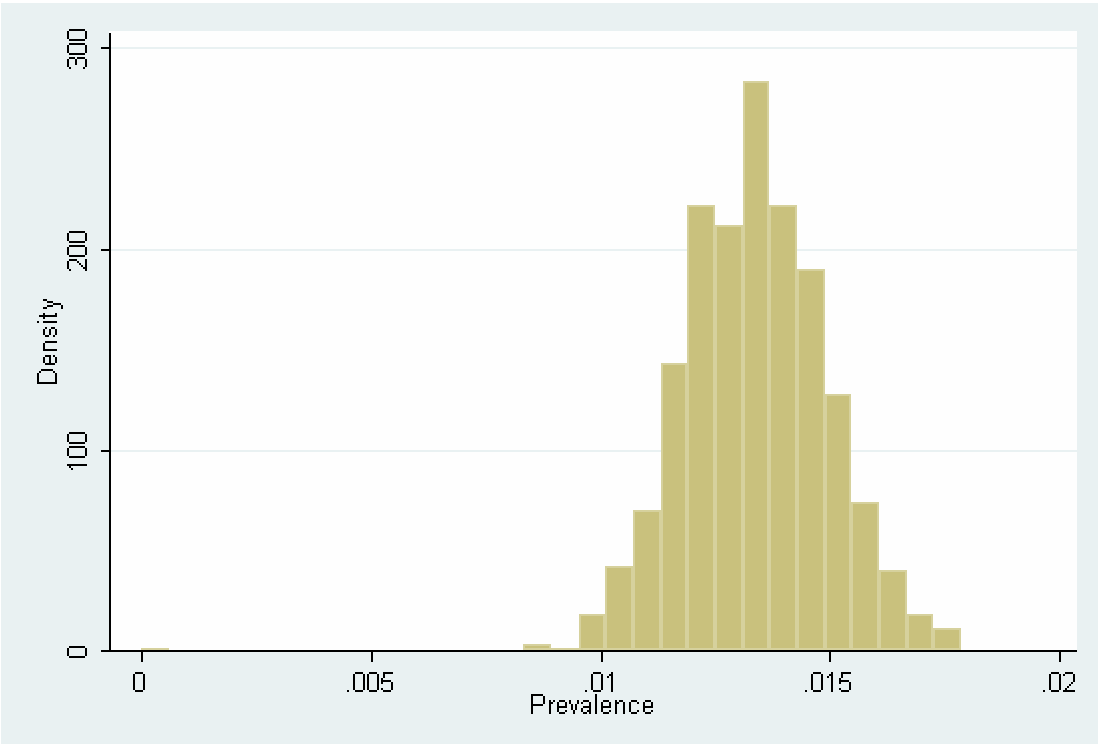


Figure 24. N=100 Simulations of Lifetime Prevalence in Women

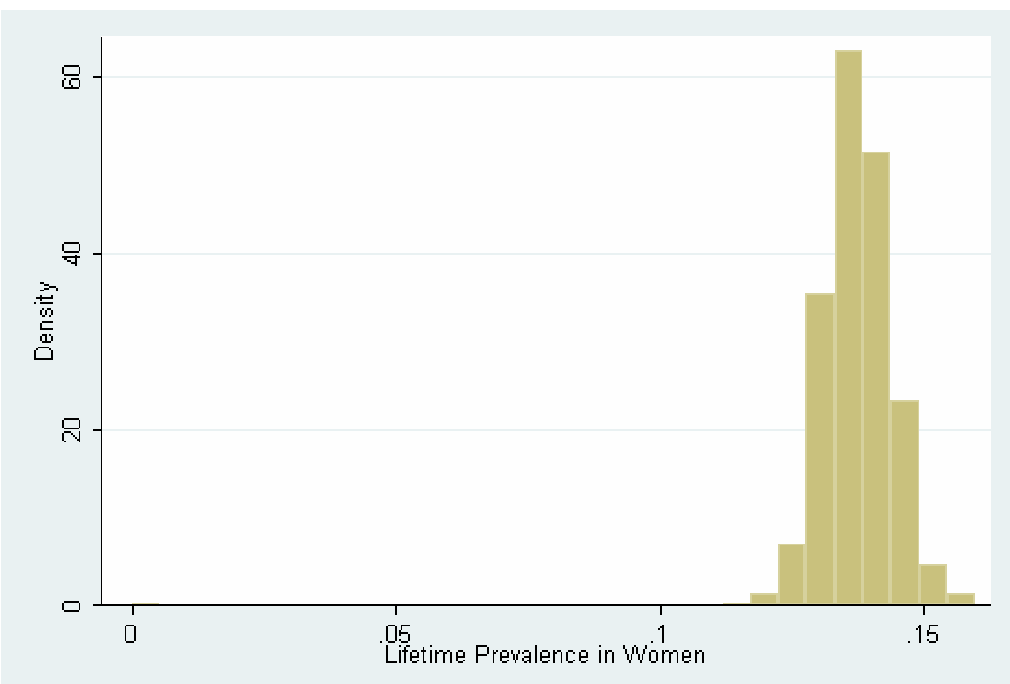


Figure 25. N=100 Simulations of Lifetime Prevalence in Men

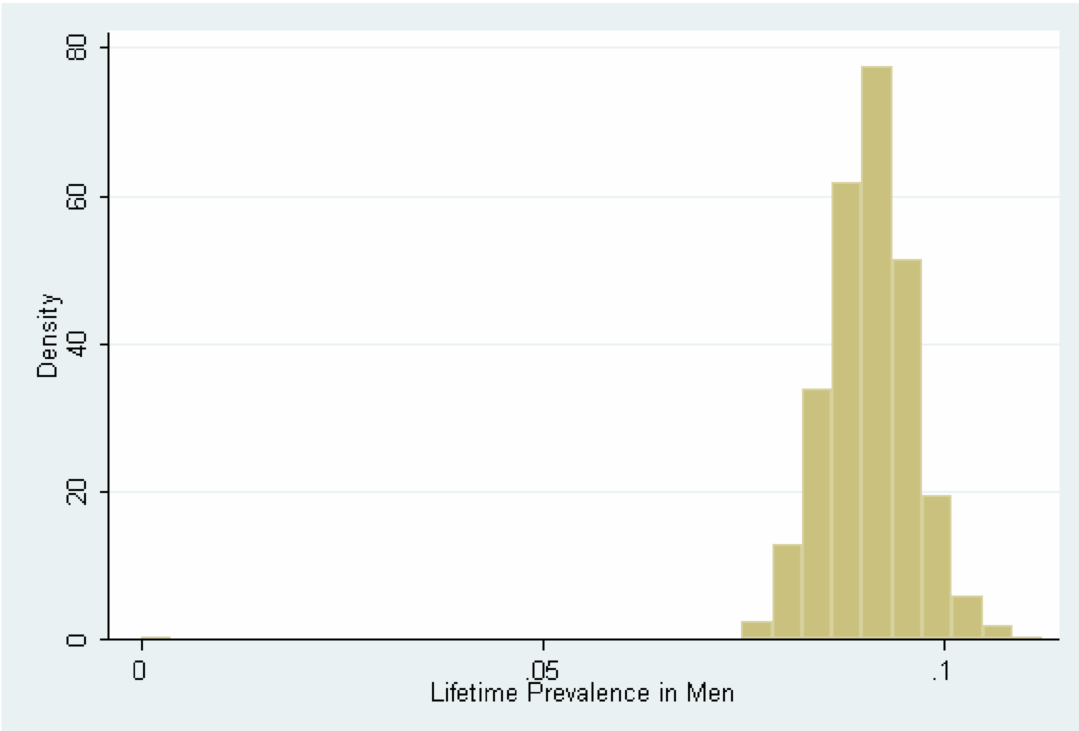
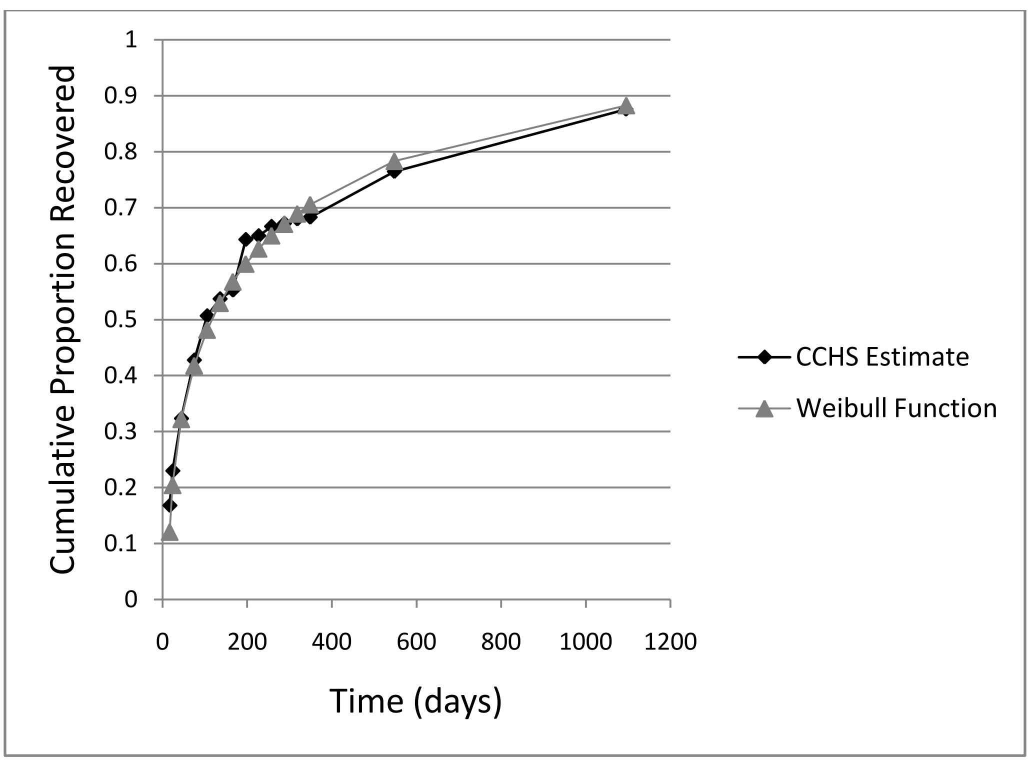


Figure 26. Observed Proportion Recovering, by Week, with the Calibrated Weibull Function*



* 14 days plus a simulated episode duration using the scale and shape parameters from Table 4.

Figure 27. Simulated and Reported Episode Durations

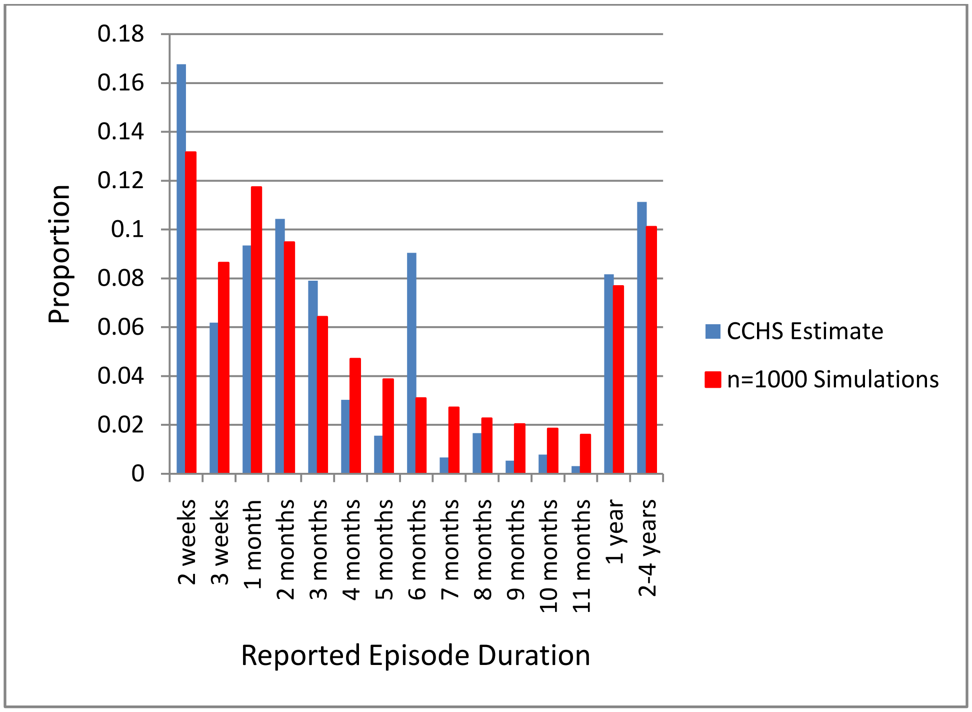
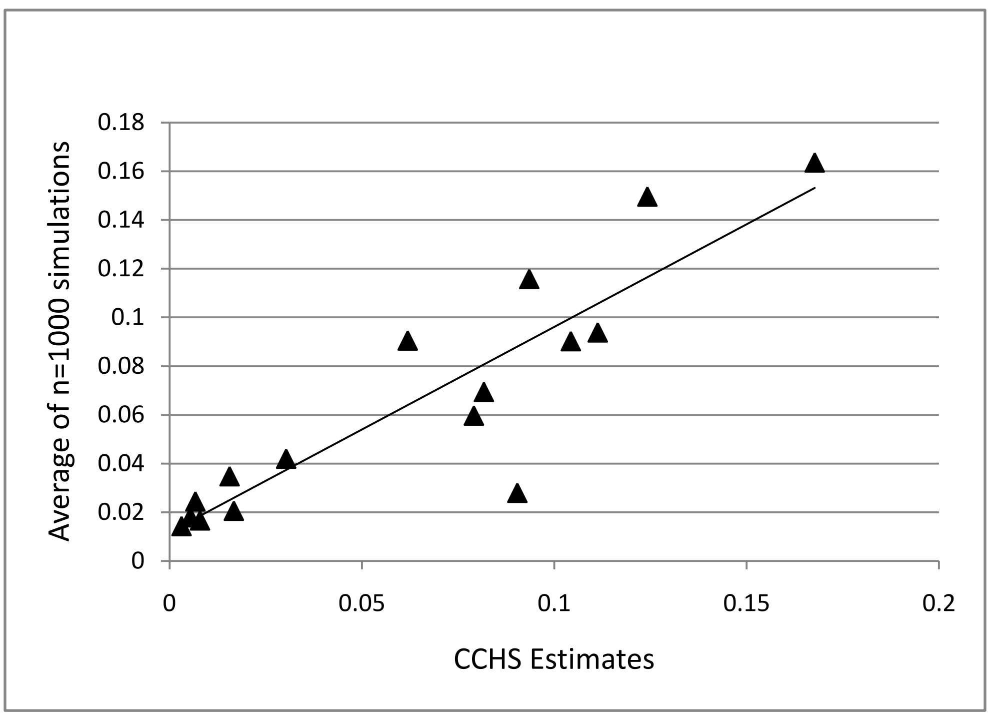


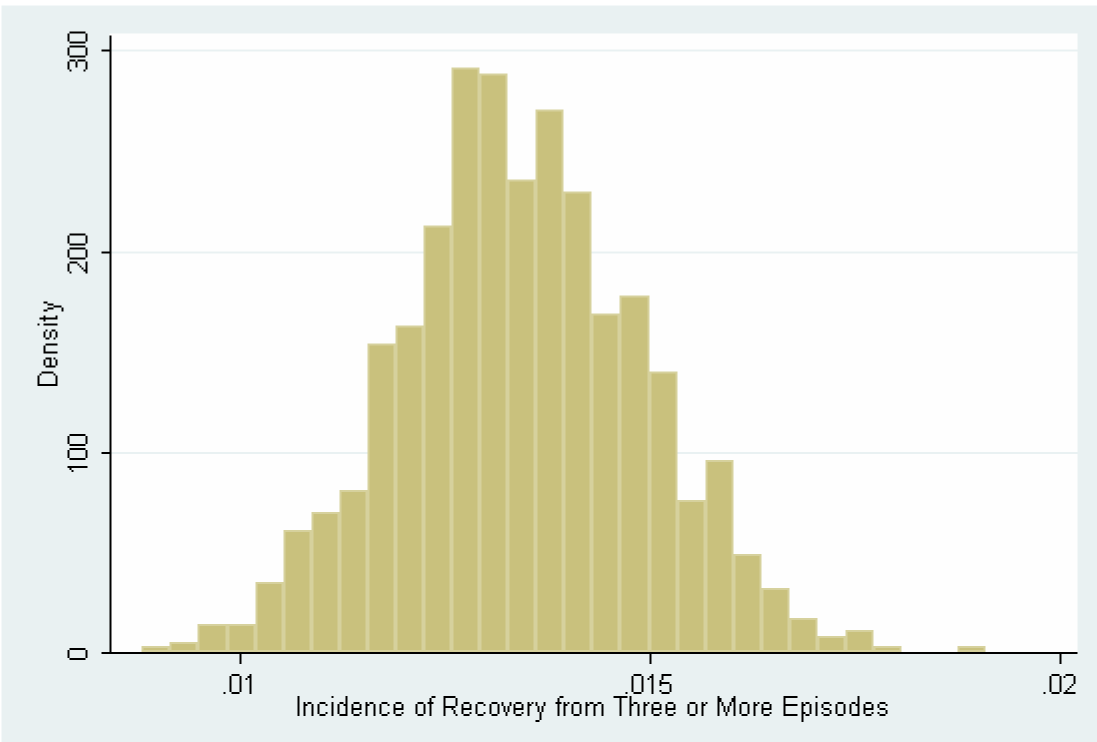
Figure 28. Quantile-Quantile Plot for Simulated and Reported Episode Duration



Part 8. Simulated 10-year Rate of Recovery from a Third Episode

Figure 29. Presents a histogram of estimates of the 10-year proportion of the population recovering from a third episode. The histogram is based on the same n=1000 simulations presented in the Figures above.

Figure 29. N=1000 Simulations: Proportion of the Population Recovering from a Third Episode in a 10-Year Period.



The estimated proportion is approximately 1.3%. Converting this to an annual rate:

$$10 \text{ year proportion} = 1 - \exp(-\text{annual rate}*10)$$

For an estimated annual rate of 13.4 per 10,000 /year.

Part 7. Sensitivity Analysis for the Effect of Mortality

Wulsin et al. (4) conducted a meta-analysis of the impact of depression on mortality. They found heterogeneity between studies but reported an overall relative risk of death for people in people with depression of 1.4. The simulations presented above were all carried out with the relative risk of death having been set to 1.0. The Wulsin estimate is a reasonable upper bound for the mortality relative risk, as most of the studies included in the review had used clinical or hospitalized patients with mental disorders. Such respondents are likely to have severe disease, so this is likely to overestimate the effect on mortality. This value was used in the modeling study by Kruijshaar et al. (3). The model was re-calibrated by OptQuest to the CCHS 1.2 data using a relative risk for mortality of 1.4. The recalibrated model continued to provide accurate simulations of prevalence (see Figure 30) as well as episode number (see Figure 31) and episode duration (see Figure 32) – with the same problematic time point at a reported duration of six months.

Figure 30. Recalibration Results for Current, Male Lifetime and Female Lifetime Prevalence, Relative Risk of Mortality = 1.4

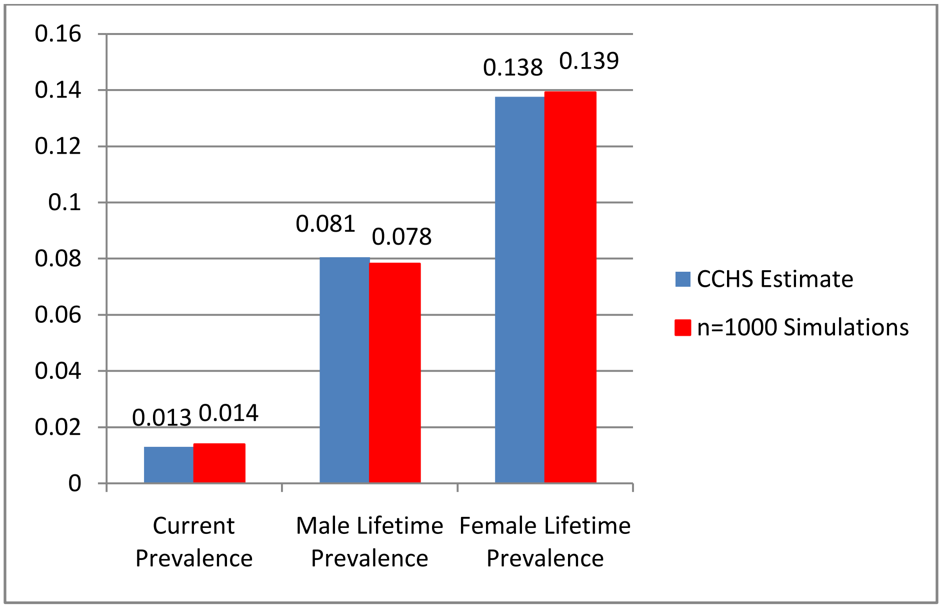


Figure 31. Recalibration Results for Episode Number, Relative Risk of Mortality = 1.4

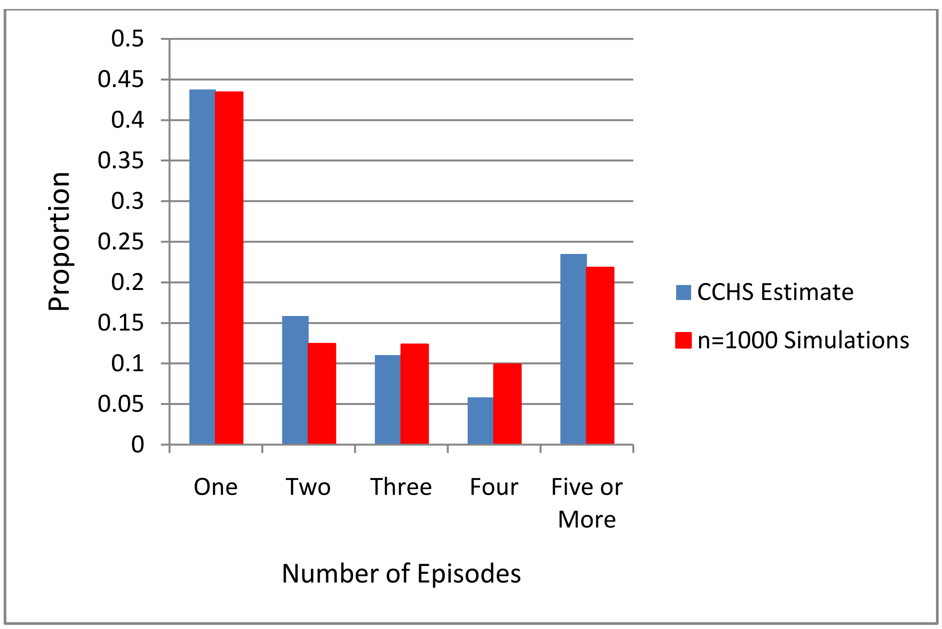


Figure 32. Recalibration Results for Episode Duration, Relative Risk of Mortality = 1.4

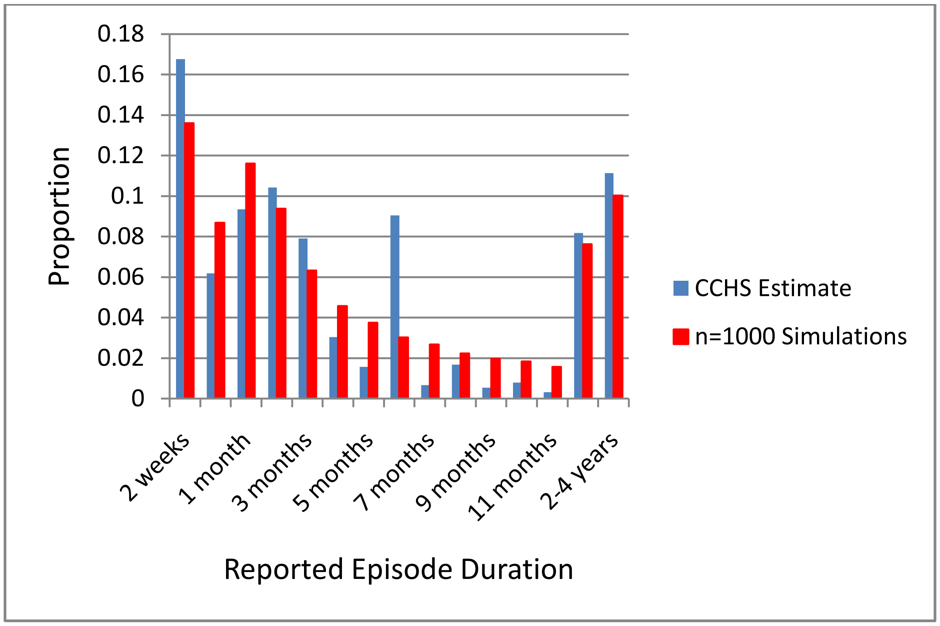
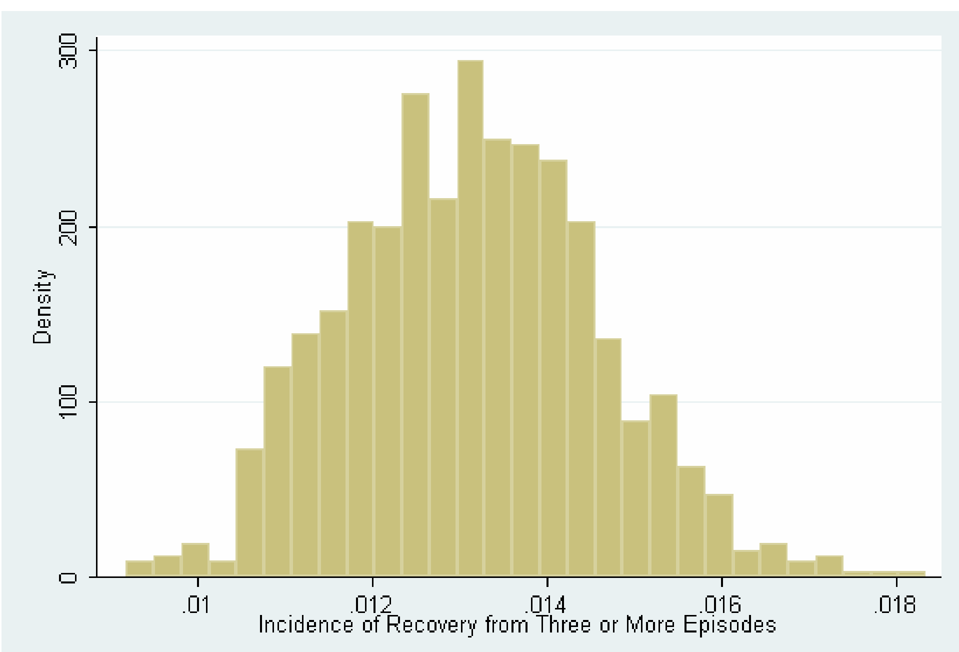


Figure 33 presents the results of n=1000 simulation runs using the model parameter values from the recalibrated model based on a relative risk of mortality of 1.4. The estimated 10-year incidence proportion for recovery from a third (or more) episode is unchanged, with the mean value of n=1000 simulations continuing to be 1.3%.

Figure 33. N=1000 Simulations of the 10-year Incidence Proportion for 3 or More Episodes from the Recalibrated Model, Relative Risk of Mortality = 1.4



Part 8. Limitations of the Simulation Model

This model provides essentially an alternative perspective on a cross-sectional dataset. Based on two types of period prevalence assessed: lifetime and 30-day prevalence, along with reported episode duration and number of episodes, a description is produced in the form of an integrated description of incidence, prevalence and mortality. The model allows the use of cross-sectional data for addressing questions that could otherwise only be addressed using a longitudinal dataset. In this instance, the goal was to estimate the rate of emergence within the population of people who may be candidates for relapse prevention using MBCT.

However, the simulations are vulnerable to problems with the validity of the underlying cross-sectional data. An important concern is the issue of recall bias. Prior reports, most notably that by Andrews et al. indicate that a sizable proportion of past episodes of depression may not be detected during subsequent interviews using the CIDI (7). Andrews postulated that the morbidity risk for major depression may be substantially underestimated by cross-sectional studies and that major depression may actually afflict most people at some point during their lives (8). Previously published simulation models have employed multiple data sources and have also found results that are consistent with the possibility that published lifetime prevalence estimates are biased downwards as a result of recall bias (3, 9). The model presented here depends entirely on retrospective recall of episodes. For this reason, it is likely that the number of recurrent episodes reported in the CCHS 1.2 is an underestimate, and therefore that the recurrence rates arising from the model calibration may also underestimate the true values. The values in Table 4 are consistent with this possibility. The average time to first recurrence according to the model is approximately 10,000 days, or 27 years. In those with multiple recurrences, the average time to recurrence is approximately 4,000 days, or approximately eleven years. A consequence is that the estimated rate of recovery from a third episode probably underestimates the true rate. However, the simulated number of past episodes reflects the CCHS 1.2 data, such that this problem is not so much a limitation of the simulation model as a potential limitation of the survey results upon which the model is based.

A weakness of the underlying data source was also evident in the episode duration data. There was a much higher than expected rate of endorsement of six months as a duration of first episodes in the CCHS 1.2. This probably reflects a choice on the part of survey participants for more accessible approximate time periods for events that are imperfectly recollected. This result serves to emphasize the tenuous nature of retrospective recall of the characteristics of past depressive episodes.

An additional limitation included the simulation of current prevalence, recurrence, multiple recurrence and recovery for men and women together. These decisions partially reflect the idea that the sex difference in prevalence is due to incidence rather than recurrence or episode duration (10) but also reflect the need to maintain simplicity in the model and to constrain its calibration to parameters for which reasonably precise sex-specific estimates could be made from the CCHS 1.2. Similarly, depiction of recurrence and multiple recurrences using a time-dependent distribution such as the Weibull distribution (rather than an exponential distribution consistent with single rates of recurrence) may ultimately provide an improved simulation, but at the expense of added complexity.

Reference List

1. Gravel R, Béland Y. The Canadian Community Health Survey: Mental Health and Wellbeing. Can J Psychiatry 2005;50:573-9.

2. Arena. [10]. 2006. Sewickley, PA, Rockwell Software Inc.

3. Kruijshaar ME, Barendregt J, Vos T, de Graaf R, Spijker J, Andrews G. Lifetime prevalence estimates of major depression: An indirect estimation method and a quantification of recall bias. Eur J Epidemiol 2005;20:103-11.

4. Wulsin LR, Vaillant GE, Wells VE. A systematic review of the mortality of depression. Psychosom Med 1999;61:6-17.

5. Patten SB, Lee RC. Refining estimates of major depression incidence and episode duration in Canada using a Monte Carlo Markov model. Med Decis Making 2004;24:351-8.

6. Patten SB. A major depression prognosis calculator based on episode duration. Clin Pract Epidemiol Mental Hlth 2006;2:13.

7. Andrews G, Anstey K, Brodaty H, Issakidis C, Luscombe G. Recall of depressive episode 25 years previously. Psychol Med 1999;29:787-91.

8. Andrews G, Poulton R, Skoog I. Lifetime risk of depression: restricted to a minority or waiting for most? Br J Psychiatry 2005;187:495-6.

9. Patten SB. A visual depiction of major depression epidemiology. BMC Psychiatry 2007;7:23.

10. Üstün TB, Kessler RC. Global burden of depressive disorders: the issue of duration. Br J Psychiatry 2002;181:181-3.