Chapter 3: Markov models

In this Chapter, we will examine one of the most common and useful ways to understand complex public health and healthcare interventions: the *Markov model.* A Markov model is a representation of health or disease that requires us to analyze multiple processes in sequence, such as multiple stages of disease, to identify how effective or cost-effective a public health or healthcare program might be. Markov models are highly flexible and allow for an infinite variety of diseases or interventions to be simulated and understood, which is why they are among the most popular tools for public health and healthcare research.

Principles of Markov modeling

You no-doubt opened this book hoping to gain skills to conquer major life-threatening diseases. Fortunately, we will now focus on a major pathology of traumatic, global significance, causing substantial suffering through severe psychological trauma: male pattern baldness (a.k.a., "MPB"). In the community of university professors, who frequently suffer from the disfiguring disorder, there are several stages of illness that illustrate the typical progression of MPB. First, the victims begin in a happy, hairy state. Next, they progress to a stage of "frontal bossing", in which a receding hairline might be ignored as it ascends up to the northern pole of the head. Third, the bossing progresses to cover the full top surface of the head, leaving a U-shaped ring of hair around the edge of the head, or—in the most horrifying case—the "comb over" hairstyle. Finally, as the

terminal state of disease, there is the complete bowling ball of comprehensive baldness. The chance of progression from one stage to the subsequent stage is about 10% each year.

Naturally, victims of this disorder do not take the problem lightly. They struggle, often in vain, to reverse the progression of MPB. Many use a hair regrowth treatment, which has a small probability of reversing the progression of disease—typically, a chance of regression from one stage to the prior stage of about 5% each year.

With this information, we might try to draw a decision tree to estimate the probability that any given person at risk for MPB might end up completely bald. Figure 3.1 provides a first attempt at this decision tree. As shown in the tree, a person starting out in the *hairy* state has a small probability per year of moving to *frontal bossing*. Once a person is in the *frontal bossing* stage, they can also try the hair regrowth treatment and have a small probability of becoming *hairy* again, or may just stay in the *frontal bossing* stage for that year, or they can progress to *combover*. This makes the decision tree complicated; if the person becomes *hairy* again, the tree could become infinitely long ....
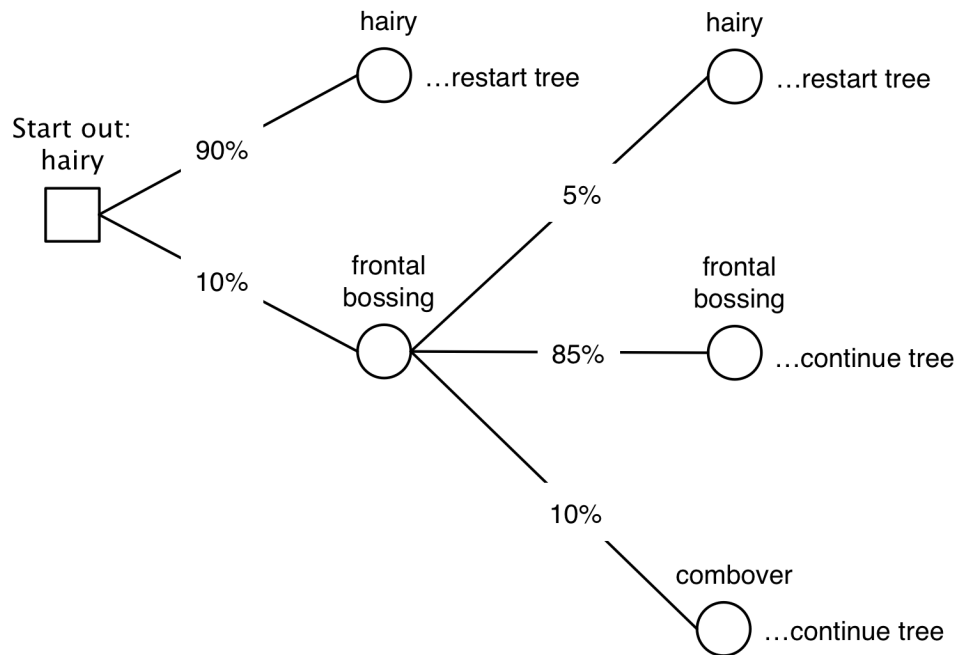
The dilemma of trying to draw an infinitely-long decision tree illustrates an inherent limitation to our approach: that diseases with multiple progression and regression possibilities can be difficult to diagram through the decision tree approach. Many diseases, such as cancers, heart diseases, inflammatory disorders, and substance abuse, are diseases that have multiple stages, with possibilities for improving disease sometimes and worsening disease at other times. These diseases are not easily diagrammed through a decision tree.

Instead of trying to draw an infinitely-long decision tree, we can use an alternative approach of drawing a *Markov model* to answer key questions about a disease, such as how likely a person is to become completely bald in the case of MPB.

A Markov model can be conceptualized as a series of lily pads on which a frog is jumping. The pads are different states of health or disease. The frog jumping across the pads is like a person going between these states of health or disease. The key features of

a Markov model are a *state diagram*, which clearly designates the possible health states that can be occupied as a sequence of connected circles; and *state transitions*, or arrows that identify the probability of moving from one state to another.

Figure 3.2 illustrates a state diagram and state transitions for male pattern baldness.
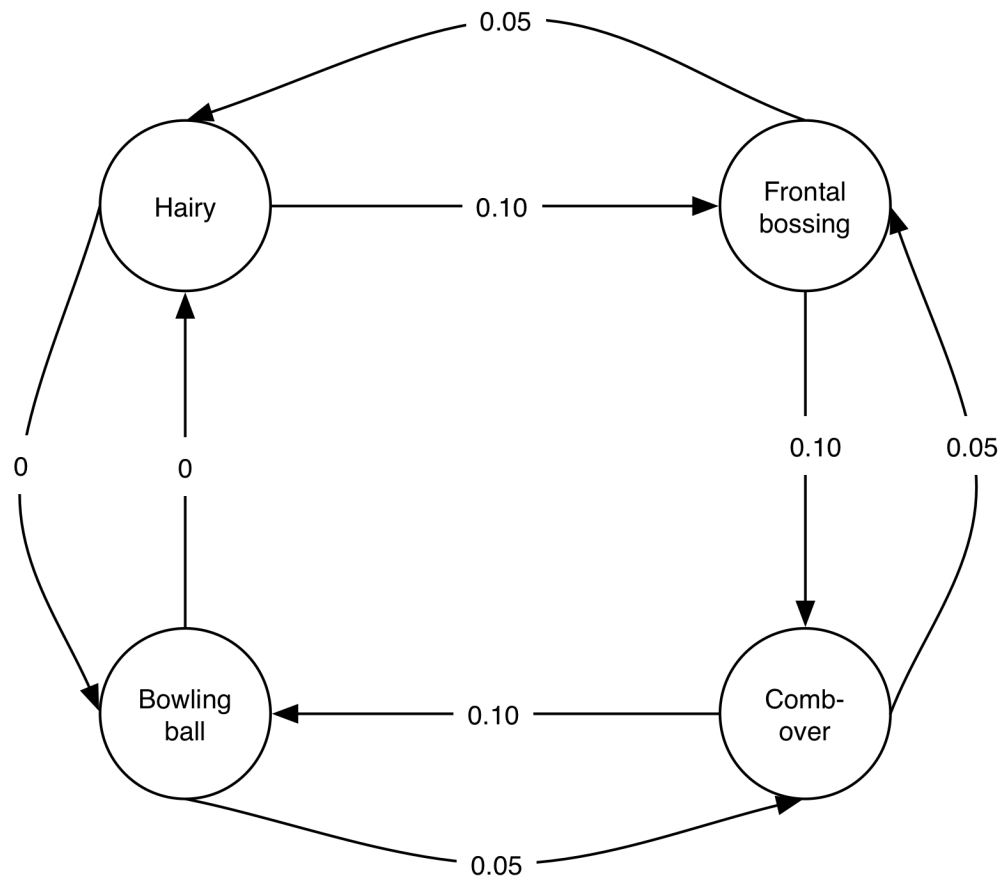
In this state diagram, we can see that there are four states of MPB: being hairy, having frontal bossing, having a combover, or being a bowling ball. The four health states are connected by state transition probabilities that describe how likely it is for a person to transition from one state to the other. It is important to ensure that we use the same

units of time across all transitions between one state and another; in this case, all state transition probabilities are in units of probability per year, also known as annual *time steps*. Markov models are said to be *irreducible* if it is possible to get to any state from any other state (even indirectly through some other state), as in this example.

Looking at the state diagram clockwise from the top left, we see that there is a probability of 0.10 per year of a *hairy* person progressing to *frontal bossing*, a probability of 0.10 per year of a person with *frontal bossing* progressing to the *combover* stage, a probability of 0.10 per year of a person with a *combover* progressing to the *bowling ball* stage, but a probability of 0 per year of a person in the *bowling ball* stage of moving straight to the *hairy* stage (if only that were possible...).

Now looking at the state diagram counter-clockwise from the bottom left, we see that there is a probability of 0.05 per year of a person in the *bowling ball* stage regressing to the *combover*, a probability of 0.05 per year of a person in the *combover* stage regressing to *frontal bossing*, a probability of 0.05 per year of a person for a person in *frontal bossing* regressing to *hairy*, but a probability of 0 per year of a person in the *hairy* stage moving to the *bowling ball* directly (i.e., they must progress through the other stages of MPB first).

Here is a key insight from drawing the Markov model as a state diagram: the diagram allows us to see how Markov modeling helps translate *individual-level* knowledge about a disease process to *population-level* insights about an overall disease. Specifically, our Markov model can help us translate data about the rates of progression and regression among stages of MPB to estimate the incidence and prevalence of balding in a population, and (as we shall later see) determine how many people might benefit from new prevention

or treatment program.

To calculate population-level estimates of incidence and prevalence using our Markov model, we need to complete our state diagram by adding all transition probabilities between states in our diagram. To do so, we can reason that our state diagram captures all states of health with respect to MPB. In other words, a person who is being described by the Markov model cannot be in a state of health that is not captured by one of the circles in our diagram. As a result, if a person does not move from one state to another, we can reason that they are remaining in the same state next year as they are in during this year (i.e., a hairy person who does not experience frontal bossing must remain a hairy person). Hence, there is a probability equal to 1 of either leaving a state to come to another state, or remaining in the same state from year to year. For example, there is a probability of 1 that a person either stays hairy, or transitions from being a hairy to having frontal bossing (there are simply no other possibilities).

Hence, we can calculate *self-loops*, or probabilities of staying in the same state each year, by calculating the probability that a person does not move to a different state but simply remains in their current state from one year to the next. The self-loops are equal to 1 minus the sum of all transition probabilities *leaving* a given state. Hence, the probability of remaining hairy next year if a person is already hairy is equal to 1 – the probability of transitioning from hairy to frontal bossing, or $1 - 0.10 = 0.90$. Similarly, the probability of remaining in the frontal bossing stage next year if a person is already in the frontal bossing stage is equal to 1 – the probability of transitioning from frontal bossing to the combover, and minus the probability of regressing from frontal bossing to the hairy stage, or $1 - 0.10 - 0.05 = 0.85$. Next, the probability of remaining in the

combover stage next year if a person is already in the combover stage is equal to 1 – the probability of transitioning from combover to the bowling stage, and minus the probability of regressing from combover to frontal bossing, or $1 - 0.10 - 0.05 = 0.85$. Finally, the probability of remaining bowling ball next year if a person is already a bowling ball is equal to 1 – the probability of regression from a bowling ball to combover, or $1 - 0.05 = 0.95$. These self-loop transition probabilities are labeled as curved arrows on Figure 3.3, which provides a complete state diagram for our Markov model. Note that the self-loops are calculated from the perspective of an individual in a given state, who is examining their probability of being in the same or different state next year. Hence, we do not count the arrows moving *into* a state when calculating self-loops; we only count the arrows *leaving from* a state when calculating the self-loops.
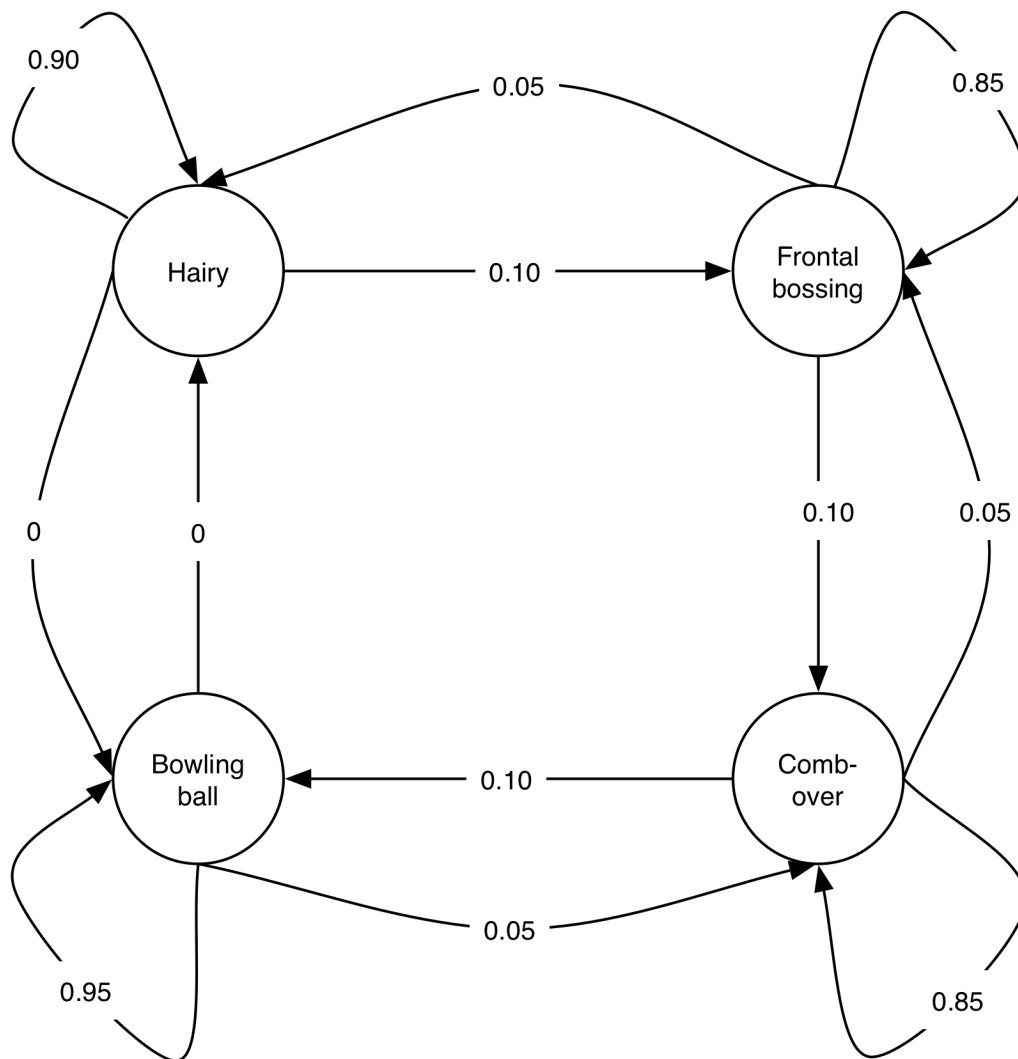
We can check that we haven't made a mistake in our model by adding all of the transition probabilities that leave a given state, including self-loop transition probabilities; the sum of all probabilities must add to 1, since people cannot disappear from our diagram.

Using Markov models to address epidemic uncertainties

Now that we have established the state diagram for our Markov model, we can use

the model to answer several critical questions about MPB. To start with, we can ask: in a male population of about 100,000 people, how many people would we expect to have a combover?

At first, this question may seem impossible to answer, since we only have limited information about the probabilities that any individual might transition between different stage of MPB, not information about the overall "epidemic" of MPB. But one of the key strengths of a Markov model is its ability to translate data from the level of the individual to the level of the population. To see this advantage of Markov modeling, we can translate our state diagram into a series of equations. Let $h\ t)$, $f(t)$, $c(t)$, and $b(t)$ refer to the probabilities that a person is hairy, in the frontal bossing stage, has a combover, or a bowling ball in any given year t. If we read our state diagram as a description of the probability that a person shifts from one state to another with each passing year, then we can write Equations 3.1 through 3.5 as follows:

[Equation 3.1] $h(t+1) = 0.90h(t) + 0.05f(t)$

[Equation 3.2] $f(t+1) = 0.10h(t) + 0.85f(t) + 0.05c(t)$

[Equation 3.3] $c(t+1) = 0.10f(t) + 0.85c(t) + 0.05b(t)$

[Equation 3.4] $b(t+1) = 0.10c(t) + 0.95b(t)$

[Equation 3.5] $h(t) + f(t) + c(t) + b(t) = 1$

Equation 3.1 states that the probability of being hairy in year $t+1$ is the probability of being hairy in year t times 0.90 (the chance of staying hairy), plus the probability of having frontal bossing in year t times 0.05 (the chance of regressing to hairiness). Note that this equation, unlike the equation for calculating self-loops, tracks all people coming

into the given state (including the chance of remaining in that state). It is a population-level equation, reflecting the overall chances of being in a given state among the entire populace at any given time.

Similarly, we can see from Figure 3.3 that Equation 3.2 is the sum of all probabilities that move a person from any state to the frontal bossing state. Analogously, Equation 3.3 is the sum of all probabilities that move a person from any state to the combover state, and Equation 3.4 is the sum of all probabilities that move a person from any state to the bowling ball state.

The final equation, Equation 3.5, tells us that people cannot magically disappear from our state diagram. A person who is potentially at risk for MBP must be in one of the states of hairy, frontal bossing, combover, or bowling ball, and therefore the sum of all four probabilities must be equal to 1.

Equations 5.1 through 5.5 are termed *difference equations* because they express the difference in the probability of being in a state at time $t+1$ and the probability of being in a state at time t.

How do we use these equations to help us understand the MBP epidemic?

The easiest strategy to utilize these equations is to first solve for the probability of being in each state under a *steady-state situation*, that is over the long-term after much time has passed and the probability of being in each state doesn't vary much over time. After we perform an algebraic solution to the problem, we will later illustrate that Markov models typically have a long-term *stationary distribution*, or a long-term steady state probability for an individual to be in any one of the given states. For now, we can take it on faith that such a long-term probability exists, and later we will demonstrate its

existence more rigorously.

Over the long-term, we can reason that the probability of being in any given state is so stable that it does not change from one time step to another. Hence, we no longer need to think of transitions from time t to time $t+1$, and can rewrite Equations 5.1 and 5.5 as their steady-state versions, Equations 5.6 through 5.10, as follows:

[Equation 3.6] $\qquad\qquad h = 0.90h + 0.05f$

[Equation 3.7] $\qquad\qquad f = 0.10h + 0.85f + 0.05c$

[Equation 3.8] $\qquad\qquad c = 0.10f + 0.85c + 0.05b$

[Equation 3.9] $\qquad\qquad b = 0.10c + 0.95b$

[Equation 3.10] $\qquad h + f + c + b = 1$

Equations 3.6 through 3.10 assume that over the long run, the epidemic of MPB will stabilize, such that the probability of being hairy will not vary dramatically from month to month, nor will the probability of having frontal bossing, a combover, or a bowling ball.

Because equations 3.6 through 3.10 provide us with four unknown terms (*h, f, c, b,* and *h*) and five equations, we can solve them algebraically. We can first solve for *h* in terms of *f* using equation 3.6, then substitute the expression for *h* into equation 3.7 to solve for *f* in terms of *c*, then substitute the expression for *c* into equation 3.8 to solve for *b* in terms of *c*, then substitute the expression for *b* into equation 3.9 or equation 3.10 to solve for the numerical value of *b* at steady state. We can then compute the other two numerical probabilities using our equations, which solve to yield $h = 0.067$, $f = 0.133$, $c = 0.267$, and $b = 0.533$. In other words, over the long-term we would expect that a minority of men (6.7%) would be hairy, slightly more (13.3%) would have frontal bossing,

even more (26.7%) would have combovers, and the greatest proportion (53.3%) would be bowling balls.

Those students who recall linear algebra can also choose to express Equations 3.6 through 3.10 as a series of matrices as shown in Equation 3.11, which can be solved with matrix multiplication to obtain the same solution.

[Equation 3.11]
$$\begin{bmatrix} 0.90 & 0.05 & 0.00 & 0.00 \\ 0.10 & 0.85 & 0.05 & 0.00 \\ 0.00 & 0.10 & 0.85 & 0.05 \\ 0.00 & 0.00 & 0.10 & 0.95 \end{bmatrix} \begin{bmatrix} h \\ f \\ c \\ b \end{bmatrix} = \begin{bmatrix} h \\ f \\ c \\ b \end{bmatrix}$$

The four-by-four matrix of transition probabilities is referred to as a *transition matrix*.

Both strategies for solving the problem are tedious, so in the next section we reveal a strategy to solve for the probabilities using the software R rather than solving the problem by hand. No matter which way we solve the problem, we should obtain the same steady-state solution.

Using our solution to the steady state equations, we can solve our problem regarding how many men we expect to have a combover in our population. Over the long-term, we anticipate that the prevalence of combovers will be the number of people in the population (say, 100,000 people in this example) multiplied by the probability that each person will have a combover, as expressed in Equation 3.12:

[Equation 3.12] $100,000 \times c = 100,000 \times 0.267 = 26,700.$

In other words, about 26,700 people would be expected to have a combover in our population over the long term. Hence, our Markov model has allowed us to estimate the *prevalence* of the disease—a population-level concept—using information about the

progression rates and regression rates between disease stages—derived from individual-level data.

Using our Markov model, we can similarly estimate the *incidence* of new disease, meaning how many people start experiencing different stages of MPB. For example, we can answer the question: what would be the typical annual incidence rate of frontal bossing? The probability per year of transitioning from hairy to frontal bossing is 10% or 0.10 per year. The absolute number of people who are hairy in a given year is $h = 0.067$, per our calculations above. Hence, the annual incidence rate of frontal bossing in our population could be expressed by Equation 3.13:

[Equation 3.13] $100,000 \times h \times 0.10 = 100,000 \times 0.067 \times 0.10 = 670.$

Hence, we expect about 670 people per year to become frontally bossed from the hairy state. Note that we did not isolate people who have never had any form of baldness before in our equation; rather, we calculated the incidence rate of frontal bossing after including all people who are currently in the hairy state. A particular individual might have progressed all the way to the combover or bowling ball stages in prior years, but has returned to the hairy state and, if they relapse to frontal bossing, would be included in our incident rate calculation.

The calculation of incidence reveals an important property of Markov models: the so-called Markov assumption, or *memoryless* property. A Markov model does not capture a person's prior history when estimating the person's probability of transitioning from one state to another. A hairy person who has never had any form of baldness before is given the same probability of having frontal bossing as someone who has previously recovered from being a bowling ball. This is unrealistic, but the fact that the Markov model lacks
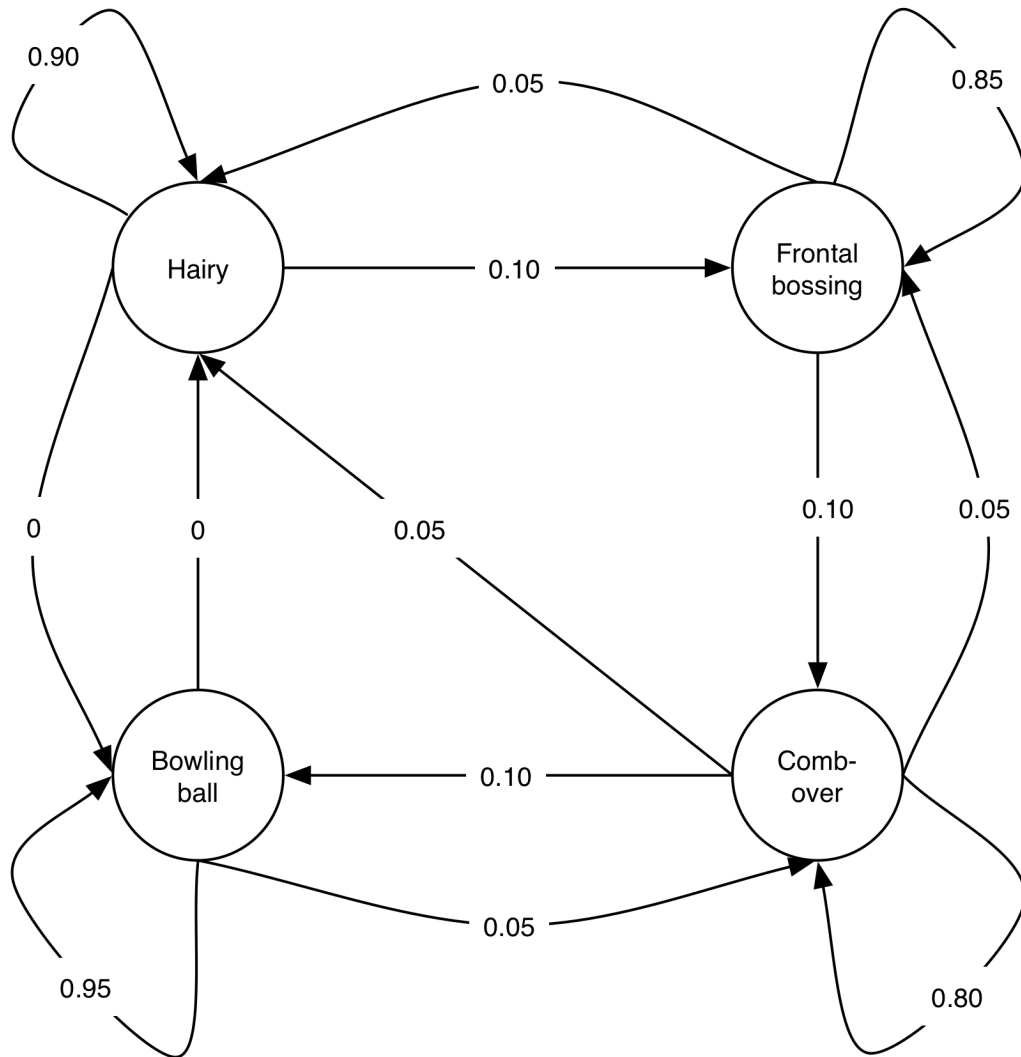
memory for a person's prior history makes the mathematics governing the model more convenient. Sometimes, modelers will add more states to a Markov model to differentiate people with a prior history of a condition from people without such a history, although this adds more states and more equations to a model, making it more complex. In a latter Chapter, we will introduce microsimulation modeling methods, which can resolve this problem. For now, let us use the simplest model for illustrative purposes.

Markov models to project the impact of an intervention

The simple Markov model we have created can assist us to project the expected impact of an intervention on population rates of disease. Suppose a new treatment has been designed that specifically helps people in the combover state of MPB to regress back directly to the hairy state. Also suppose an emergency relief fund is being established to help the "victims" of combover (meaning the people affected, not those who have to look at the combovers... though the latter may sometimes be considered victims too). We know, from our calculation above, that the relief fund would need to aim to treat the estimated 26,700 people currently with combovers in our population.

Suppose the new treatment for combovers—a spray-on solution that helps stimulate hair follicles in the middle of the head—helps people transition from the combover state to the hairy state at a rate of 5% or 0.05 per year (as opposed to the current rate of 0%, because there is no current rate of direct regression from combover to hairy). The spray-on solution is not a permanent fix; people who return from combover

to hairy are still subject to transitioning to frontal bossing just like everyone else in the hairy state. Figure 3.4 illustrates how the new treatment would change the diagram for the Markov model:

0.90   0.05   0.85

Hairy   0.10   Frontal bossing

0   0   0.05

0.10   0.05

Bowling ball   0.10   Comb-over

0.05

0.95   0.80

Note, importantly, that if we simply added an arrow to connect the combover state to the hairy state, we would experience a big problem: the sum of probabilities from the combover state would add up to more than 1, implying that people appear out of nowhere

and multiply from the combover state (ack, combovers reproducing themselves!). We instead envision that the increased probability of recovering from the combover state to the hairy state due to the new spray-on hair treatment would reduce the self-loop probability of staying in the combover state, from 0.85 to 0.80 per year (so that the total probability of staying in the combover state or moving to any other state still sums to a value of 1).

We can depict the change in our model following the introduction of the new spray-on hair intervention by rewriting Equations 3.6 through 3.10 as Equations 3.14 through 3.18:

[Equation 3.14] $\qquad h = 0.90h + 0.05f + \mathbf{0.05c}$

[Equation 3.15] $\qquad f = 0.10h + 0.85f + 0.05c$

[Equation 3.16] $\qquad c = 0.10f + \mathbf{0.80c} + 0.05b$

[Equation 3.17] $\qquad b = 0.10c + 0.95b$

[Equation 3.18] $\qquad h + f + c + b = 1$

The two changes capturing the impact of our intervention are reflected in bold: the addition of $0.05c$ in Equation 3.14, and a change from $0.85c$ to $0.80c$ in Equation 3.16. No other equations are directly changed due to the intervention, although we will see in a moment that the probability of being in any state, and subsequent estimates of prevalence and incidence. In other words, even though only two states (combover and hairy) have been directly affected, the Markov model will help us account for the complex interconnected ways that other states will be affected as well through indirect means.

When we solve Equations 3.14 through 3.18 (either algebraically or using matrix methods, as described above) we see a subtle change in the solution for the stationary

distribution transition probabilities. Before our spray-on treatment intervention, Equations 3.6 through 3.10 solve to yield $h = 0.067$, $f = 0.133$, $c = 0.267$, and $b = 0.533$. After our spray-on treatment intervention, Equations 3.14 through 3.18 solve to yield $h = 0.200$, $f = 0.200$ $c = 0.200$, and $b = 0.400$. Hence, the probability of having a combover, because of our intervention, reduced from 0.012 to 0.009 over the long-term. Overall, the prevalence of combovers would be expected to lower from 26,700 people before our intervention to:

[Equation 3.19] $100,000 \times c = 100,000 \times 0.200 = 20,000.$

Hence, our intervention would help lower the prevalence of combovers by 6,700 people in our population.

In addition to calculating the lower *prevalence* of combovers, we might also ask whether our intervention has indirect benefits in terms of altering the *incidence* of frontal bossing. The absolute number of people who are hairy in a given year is now $h = 0.200$ after the intervention. Hence, the incidence in the population per year of frontal bossing would be expressed by Equation 3.20:

[Equation 3.20] $100,000 \times h \times 0.10 = 100,000 \times 0.200 \times 0.10 = 2,000.$

While our previous incidence per year was 670, we now have a higher incidence of frontal bossing of 2,000 because the overall population of hairy people is now higher (more people with combovers have become hairy again) but the probability of transitioning from hairy to frontal bossing hasn't changed. This doesn't indicate that our intervention is "bad thing", because we have still lowered the overall probability of having a combover (from 26.7% to 20.0% per the stationary distribution solution above), and also of becoming

a bowling ball (from 53.3% to 40.0% per the stationary distribution above). The result simply highlights that the outcomes of Markov models can be complex, and we should look carefully at how indirect effects between states can lead to surprising conclusions.

How might our intervention affect the disease in the short-term, not just in the long-term steady state? And what would be the incremental cost-effectiveness of the intervention at a given price? In the next Chapter, we will program our model into R to estimate the cost-effectiveness of the intervention and realize the potential of using a Markov model to estimate the population-level benefits of a health intervention.