Speaker's notes

Dia4:

The concrete objective of our study was construction of a model of the dynamics of functional health based on National Population Health Survey data. This was accomplished by modeling each component attribute of the Health Utility Index: a generic index of functional health status. The HUI is based on eight attributes: vision, hearing, speech, mobility, dexterity, cognition, emotion, and pain.

Want to emphasize here that we are concerned with FUNCTIONAL HEALTH STATUS – that is with degree of proficiency / functioning on the 8 attribute dimensions. As such, our intuitions regarding, for example, underlying disease states or chronic conditions may not apply here; unless a disease or chronic condition is closely linked with proficiency / functioning .

The degree of proficiency has been measured using sets of simple questions that, for example, determine whether a respondent requires eye-glasses or lenses to correct his/her vision.

Finally, Functional Health status is an essential component of Health Adjusted Life Expectancy – a key summary health indicator and one whose dynamics and distribution is the fundamental objective of this study.

Dia5:

n.b. this is more ambitious than usual statistical analysis where the "dependent variable" is a scalar. In this analysis, the dependent variable is a set of relationships – a representative sample of cohort health-adjusted life lengths (HALLs) underlying (microanalytically, not Sullivan method) HALE = Health-Adjusted LE

Dia6:

This diagram illustrates calc's for both conventional life expectancy, LE, and the extension to HALE. Conceptually, HALE = generalization of a conventional survival curve; life expectancy is simply the area under the curve. But this standard LE calculation implicitly treats every person-year of life of this birth cohort as exactly identical. i.e. takes no account of variations in health among the living. These variations are represented in the diagram by the various shadings of gray beneath the survival curve -- a darker shade implying a worse health state.

Could define a threshold level of health below which individuals ≡ "not healthy", or as disabled, and then compute the area under this lower survival curve, shown in blue – often called "disability-free" life expectancy (DFLE).

But dichotomization of individuals' health status is too rough. Better approach is to compute the area under the survival curve <u>weighted</u> by the lightness or darkness of the shading. A year of life in full health counts as a 1, while a year of life in much poorer health would count as a 0.3, say. Result of this

weighted calculation is HALE – most appropriate summary measure for assessing progress in population health.

n.b. all of this so far can be done with semi-aggregate / life table methods; but much better to do microanalytically – using individual HALLs and LLs, which is what the HealthPaths model, to be described, does

Dia7:

The health of the youngest (under 10), as assessed by the HUI, is characterized by perfect or near perfect health. At succeeding ages, the proportion at or near perfect health declines and the range of HUI over which the remainder of the population is distributed increases. These basic facts can be seen in empirical distribution functions for each of the 10-year age groups.

HUI appears to provide a highly plausible description of the affect of aging on population health. However, our concern here will be with the harder problem of providing a realistic description of individual health trajectories / dynamic life cycle profiles in terms of an appropriate set of equations and then going beyond that to determine the general properties of those equations using microsimulation.

Dia8:

We use the ordinal character of each HUI attribute together with the continuous time simulation facility of ModGen – our software environment - to construct a parsimonious model: 16 equations rather than 240. This simplification relies on an interpretation of differences in attribute states between interviews as reflecting a process of incremental change. It is the two year interval between interviews that accounts for the comparisons of vision involving observations at t and at t+2.

Vision Ordered Categories

1. *No visual problem*
2. *Problem corrected by lenses*
3. *Problem seeing distance - not corrected*
4. *Problem seeing close - not corrected*
5. *Problem seeing close and distance - not corrected*
6. *No sight at all*

Dia9:

The explanatory component of our model was built up from a selection of proxies that – hopefully – cover an appropriate range of influences on health.

Education, Daily Smoking (including cumulative smoking years) and BMI are familiar covariates.

The psycho-social factor 'Sense of Coherence' is less familiar. It is a standardized scale that attempts to measure an individual's capacity to cope with stress.

Most social scientists would never use the term 'proxy'; instead they use terms like 'variable' or 'construct'. But, we use 'proxy' in order to emphasize that there is undoubtedly much more underlying complexity than can be captured in any given small set of variables.

Dia10:

The relations that constitute our model are represented here over 3 cycles of the NPHS. There are two types of relations acting over a given span of time:

Intrinsic relations: where incremental change in any HUI attribute may be influenced by any other attribute or lagged attribute.

Extrinsic relations: where incremental change could also be influenced by Age, Education, Smoking, BMI, & Coherence. Change in each of these extrinsic factors are themselves influenced by the past – including feedback from lagged HUI.

Finally, the relation between mortality and health is expressed as an age-sex-specific relative risk as a function of HUI. Simulated HUI can be derived (using the weighting algorithm) from its underlying simulated attributes / dimensions.

In the equations specifying attribute change hazards, these relations imply a complex set of candidate covariates:

- lagged and cross-lagged effects of each attribute level on chances of attribute change implies 80 coefficients per equation;
- extrinsic effects and interactions with age add an additional 174 coefficients per equation;
- interactions of all those covariates with sex and three broad education levels add a further 804 coefficients per equation

This brings the candidate total for each of the 16 hazard equations to 1072 coefficientsDia:

Dia11:

These hazard regressions have been estimated using the Forward Stagewise Regression Algorithm, which is astonishingly simple. It consists of an iterative process, a succession of small steps in each of which the covariate most strongly associated with the working residuals is identified and its coefficient is incremented (or decremented) by a small amount.

Because the algorithm only requires inspection of a vector of association measures, the number of covariates is limited only by the memory or data storage available. It can easily handle the 1072 candidate covariates included in the hazard equations for change in HUI attributes.

In addition to accommodating large numbers of candidate covariates (possibly many more than the available sample size), the Stagewise Algorithm effectively chooses a model. That is because, using an appropriate termination rule, many coefficients will be left at the initial value of zero. The set of non-zero coefficients represent the selected model, which may vary from sample to sample.

We have used a simple Poisson likelihood for this stage of estimation.

Dia13:

In working with our model, bootstrapping is made to play three roles:

First, bootstrap weights provide a convenient partition of the sample into two parts: one part that can be used for estimation and one part that can be used to prevent over-fitting by monitoring out-of-sample prediction error. In general, just over 1/3 of PSU's that are left out of a given bootstrap.

Second,in this application, bootstrapping provides an estimate of coefficient variance that includes model selection variance.

Third, bootstrap simulation will be used (beyond the estimation process) to quantify variance in simulation outputs by means of replicate simulations based on equations estimated from separate bootstrap samples.

Dia14:

Here's an illustration of our use of the Forward Stagewise Algorithm. This illustration uses a special small version of the Worse Vision Equation. As, there were only about 300 terms in this illustrative equation, it was possible to estimate the full unrestricted form of the equation using SAS Proc GENMOD.

The chart displays (for one bootstrap sample) the difference between the Out-of-Sample Error from the Forward Stagewise Regression and the Out-of-Sample Error from the corresponding Full Equation.

Out-of-Sample Error corresponds to the Poisson Deviance calculated using the portion of the data that was not included in this particular bootstrap sample, i.e. had zero weight.

The difference between errors is positive when the Forward Stagewise Regression is inferior and negative when the Full Equation is inferior.

It took about 350 steps to find a model that had minimal prediction error; and this model was clearly superior to the Full Equation.

In estimating the hazard equations that were used in the simulations, we terminated the Forward Stagewise algorithm as soon as the prediction error began to increase.

Then, as a final step, a log-normal person-specific error variance was estimated, holding the Stagewise coefficients fixed.

Dia15:

The process of estimating an equation using the Forward Stagewise Algorithm to minimize prediction error can be repeated with many bootstrap samples.

The result is multiple sets of estimates that together display the combined effects of variance in estimation and of model selection uncertainty. This represents a significantly more general evaluation of total uncertainty than is usually available.

This strategy – replicating model selection with different bootstrap samples - moves us closer to understanding the "total error" in statistical analysis: an as yet little recognized benefit of the bootstrap weights that Stat Can has been releasing with an increasing number of its surveys.

Dia16:

In the applied microsimulation world, the usual approach is to plug-in the best parameter estimates, and then to simulate the implied outputs. Often 'calibration' or 'alignment' is then needed to ensure that the results can pass basic validity tests; for example, approximating historical benchmarks given appropriate settings.

Our use of the same set of bootstrap samples for estimation of each equation (both Attribute Change and Covariate equations) permitted a novel approach to simulation -- what might be called Bootstrap Simulation.

We use many bootstrap estimates of our parameters to produce many simulated outputs, which leads to an direct estimate of the expected value of that output.

Since simulated outputs will often be a non-linear function of the parameters, the expected value of the outputs will generally not be simulated using the conventional approach.

Dia18:

As evident from the description of the NPHS analysis, the data show important connections among several of the variables. As a result, a rigorous scientific approach requires that we need to think in terms of a "complex web of causality", as emphasized by Nancy Krieger.

This diagram shows one way to visualize such a complex web. The grid of "blobs" represents an individual's stylized biography

n.b. health status shown as rectangle since it is an index based on 8 underlying dimensions; circles represent single variables

The causal web is indicated by the many arrows connecting different blobs, with only a few of all the possible arrows shown (only those of first order, i.e. no 2+ period lags), and no indication is given of the quantitative strength along any of the arrows .

One of the holy grails of longitudinal surveys is precisely to estimate which of these arrows is significant, and then to quantify the strength of the specific pathway. But in this HealthPaths analysis, our objective is much more ambitious – we seek to quantify the entire network of relationships indicated in this diagram, and then draw out their joint implications

Dia19:

The HealthPaths microsim model is written in Stat Can's ModGen, and draws inspiration from many years of experience with two other models – POHEM and LifePaths.

Simulation goes one individual at a time, with the birth of a synthetic individual. The individual is aged by being exposed to the health transition patterns of the Canadian population, whose estimation from the NPHS has just been described.

n.b. shows 2nd order Markov – this is a compromise: better than 1st order, but higher order likely also significant; on the other hand, with 2nd order, still have enough longitudinal data to estimate individual effects – i.e. represent individual heterogeneity, which turns out to be significant

n.b. this image gives impression of discrete time, but in fact we use continuous time, since transitions (among discrete states) are modeled in terms of waiting time distributions

Dia20:

Process is repeated many times to create a large sample – 2 million in simulations to be shown – of such synthetic individuals (replicants?)

Simulation sample size is made large enough to ensure that Monte Carlo error, from the use of random numbers, is small relative to the model outputs of interest.

Dia21:

HealthPaths works with overlapping birth cohorts embedded in real calendar time.

However, for results to be shown, focus is on 1960 birth cohort – n.b. empirical dynamics do not have time subscript; i.e. patterns of dynamics (hazard rates) are assumed constant over time

First step will be to simulate our base case scenario for this cohort

Dia22:

Once we have a base case scenario that is our best estimate of the history of the 1960 birth cohort, and our best judgment of this cohort's projected future, then our virtual (in silico) lab apparatus is ready for use. Key method we are using is to "knock out" one or a group of the empirically observed connections, re-run the simulation, and see what difference it makes. This approach is very much akin to the knockout mouse models used in genetic research. But I think it is very novel in social and health science research. The power of microsimulation is that it gives us the required tool.

As noted in this diagram, this analysis is still in part illustrative, a proof of concept. So we have chosen to look at only a few key variables (a spider with rather limited vision, in Nancy Krieger's terms).

Health status is very richly characterized in the NPHS using the McMaster HUI, with its underlying 8 dimensions; smoking and obesity (measured by BMI – n.b. self-reported) are obvious factors.

There is also a great deal of evidence re SES factors; we've chosen educational attainment because, e.g. compared to income, it saved us quite a bit of work as we could avoid modeling income as a co-evolving individual characteristic.

Dia23:

Finally, as a bit of a flyer, we included Antonovsky's Sense of Coherence scale. This is likely not very well known, but was included on the NPHS as an exploratory variable. It has not been used much by researchers analysing the NPHS, and we thought it would be good, since our decision in 1993 to include it in the NPHS, to see if it was at all important in this kind of analysis.

In sum, the SoC is designed to measure the extent to which individuals feel their life is comprehensible, manageable, and meaningful. The quote is from Antonovsky himself, who died prematurely in the 1990s, just as the health promotion community was beginning to embrace his concept of salutogenesis – the factors driving good health, rather than those driving disease.

Dia24:

So here is the first of our results, for the base case scenario. Overall, HALE for men in this cohort is projected at 66.5 years, and for women 71.9.

Because of the very sophisticated way the network of equations was estimated using the set of bootstrap weights, we have run each simulation 40 times in order to have direct estimates of the bootstrap uncertainty. Let me emphasize that this uncertainty is far more inclusive than the usual kind. Monte carlo error is here, but with 2 million cases, it has been reduced to 0.2. or 0.3 years.

The sampling error of the regression coefficients is here. But in addition we have explicitly taken account of any correlated errors across the estimated equations, as well as specification errors based on the iterations on out-of-sample predictive accuracy.

The rest of the graph shows the 9 decile cut points in the distribution of HALLs for men and women

n.b. turning on its side, like a survival curve, but of HALLs rather than LLs (recall earlier graph) – so extent of variation not at all surprising

n.b. also, this is NOT our preferred idea of health inequality, in contrast to Julian LeGrand and Chris Murray et al

Dia25:

While it is trickier to do, this graph is our best attempt to show the <u>joint distribution</u> of LLs and HALLs

e.g. the longest line to the right is for individuals who lived to age 100+. The dist'n it shows is for the HALLs of this long lived sub-population – e.g. most (over 3%) of those with LLs of at least 90 lived with HALLs about 10 years less than their LLs.

If we focus on age 60, or 65, and go up a vertical grid line, we see that the proportions with HALLs at these ages increases as the LLs fall – showing that those who are in poorer health at age 60 or 65 have shorter LLs on average.

And for those with LLs up to age 35, the graph shows that virtually all their HALLs = their LLs, i.e. they essentially experience sudden death.

Dia26:

This graph explores the relative importance of each of the 8 dimensions of health status underlying the summary health index.

Basically, each dimension was set to its best "full health" level, one at a time, and the base case simulation was rerun.

Doing these experiments with speech and dexterity had virtually no effect, indicating that problems with these dimensions of health status were either not very prevalent, or not given much weight in the tariff or valuation function used to do the aggregation, or both.

On the other hand, cognition was most important, followed fairly closely by pain and emotion for men, and pain, emotion and vision for women.

Hypothetically eliminating the effects of cognitive decline for the 1960 birth cohort would increase HALE by over 5 years for men, and over 4½ years for women.

Dia27:

By far the most important factor associated with health status dynamics is chronological age This is illustrated here with a trio of simulation experiments where we hypothetically set individuals to have the transition dynamics of an always 20 year old, an always 50 year old, and an always 80 year old.

The base case scenario is close to that of the always 50 year olds.

Dropping age to always 20 adds about 5 years to HALLs, while increasing age to always 80 has a much more dramatic effect – reducing HALLs by up to 40 years.

Interestingly, the always 80 year old scenario results in a much wider spread of HALLs across the deciles. The intuition here is that at higher ages, the extent of transitions among health states is greater.

Recall that all these simulations include information on 2nd order Markov, i.e. two period lagged, relationships. Essentially, at ages 50 and younger, there is much less likelihood of a deterioration in health status, so individuals' HALLs "smear out" to a much lesser extent.

Dia28:

There is ubiquitous evidence of the important role of SES as a determinant of health. So we were keen to do a "knock out" experiment using our SES covariate, educational attainment.

This graphs shows the hypothetical effects as if everyone either never completed high school (Low Ed), or everyone not only completed high school, but also a first university degree as well (High Ed). The base case is shown by the dashed lines.

The effects are not as great as those associated with age, but they are considerable. They are also asymmetric

– an improvement in educational attainment has a much smaller effect than a reduction, on the order of one versus five years; and

– the effects of educational attainment are greater in the lower deciles of HALL

Dia29:

There is ubiquitous evidence of the important role of SES as a determinant of health. So we were keen to do a "knock out" experiment using our SES covariate, educational attainment.

This graphs shows the hypothetical effects as if everyone either never completed high school (Low Ed), or everyone not only completed high school, but also a first university degree as well (High Ed). The base case is shown by the dashed lines.

The effects are not as great as those associated with age, but they are considerable. They are also asymmetric

– an improvement in educational attainment has a much smaller effect than a reduction, on the order of one versus five years; and

– the effects of educational attainment are greater in the lower deciles of HALL

Dia30:

Here they are, men at the top for the 4 scenarios, women at the bottom.

First impressions are that smoking has the largest impact, but remember there are only two scenarios in this case – status quo versus everyone always a never smoker. So the more appropriate comparison is the range from low to high HALE

Dia31:

The "braces" here indicate these more relevant ranges. We were certainly surprised to see that SoC has the greatest impact. This finding merits further exploration. It does fit with the earlier finding that among the 8 dimensions of health status underlying these results, cognition was most important.

We were not as surprised to see the smallest impact for BMI, notwithstanding all the public discussion of the "obesity epidemic". Other studies from the NPHS as well as NHANES (esp. Flegel et al.) have already shown that the RR curve actually has a minimum in the "overweight range" between 25 and 30 rather than in the "normal weight" range of 20 – 25.

Education and smoking are in between in their impacts – accounting for …

Dia32:

Let me conclude with the general caveat that these results are brand new and should be taken as exploratory and tentative.

They do demonstrate not only the feasibility but also the power of the novel statistical methods we have used for estimating our coherent network of health status and co-evolving risk factor dynamics – drawing on the fortuitous set of bootstrap weights that are part of the NPHS (and CCHS).

The results also demonstrate the benefits of a close coupling of the program of empirical estimation and model construction.

HealthPaths illustrates an extension of microsimulation methods for epidemiology and health policy potentially of the same order of importance as the knockout mouse model in genetics.

And it illustrates the feasibility and salience of HALE and its related family of health indicators.

The substantive results are certainly provocative – the importance of smoking and education is in line with previous research, but the relative weakness of BMI and the unexpected strength of the role of SoC are perhaps surprising.

Finally another caveat concerns the sampling variability of these results – it is quite high, but it must be remembered that it includes far more sources of variation than is typical in these kinds of analysis.

Thank you