© Health Research and Educational Trust
DOI: 10.1111/1475-6773.12044
METHODS

Using Common Random Numbers in Health Care Cost-Effectiveness Simulation Modeling

*Daniel R. Murphy, Robert W. Klein, Lee J. Smolen,
Timothy M. Klein, and Stephen D. Roberts*

Objectives. To identify the problem of separating statistical noise from treatment effects in health outcomes modeling and analysis. To demonstrate the implementation of one technique, common random numbers (CRNs), and to illustrate the value of CRNs to assess costs and outcomes under uncertainty.

Methods. A microsimulation model was designed to evaluate osteoporosis treatment, estimating cost and utility measures for patient cohorts at high risk of osteoporosis-related fractures. Incremental cost-effectiveness ratios (ICERs) were estimated using a full implementation of CRNs, a partial implementation of CRNs, and no CRNs. A modification to traditional probabilistic sensitivity analysis (PSA) was used to determine how variance reduction can impact a decision maker's view of treatment efficacy and costs.

Results. The full use of CRNs provided a 93.6 percent reduction in variance compared to simulations not using the technique. The use of partial CRNs provided a 5.6 percent reduction. The PSA results using full CRNs demonstrated a substantially tighter range of cost-benefit outcomes for teriparatide usage than the cost-benefits generated without the technique.

Conclusions. CRNs provide substantial variance reduction for cost-effectiveness studies. By reducing variability not associated with the treatment being evaluated, CRNs provide a better understanding of treatment effects and risks.

Key Words. Variance reduction, common random numbers, cost-effectiveness modeling, osteoporosis

Modeling in Cost-Effectiveness Analysis

Cost-effectiveness (CE) models have been used as tools in virtually every health policy decision for several years. Models, by nature, are mathematical/logical reductions (simplifications) of reality. The quality of a model is a function of three features:

1. How the decision problem is represented by the model,
2. The quality of data used within the model, and
3. How accurately the model represents the critical causal interactions between resources and treatments, treated patients, and treatment outcomes.

Several types of health care CE models are available to the analyst and decision maker: decision trees, Markov models, discrete event simulation models, and disease transmission models among others. The objective of any model is to organize and accurately represent the major causal factors that impact the decision process being addressed.

For the reader wishing to learn more about health care modeling in policy analysis, recent literature contains guidelines on good modeling practice (NICE 2008; Caro, Briggs, Kuntz, and Siebert 2012), model selection (Barton, Bryan, and Robinson 2004; Brennan, Chick, and Davies 2006), and the use of modeling in the drug development process (Boyd, Fenwick, and Briggs 2010; Burman and Wiklund 2011; Petrou and Gray 2011).

A prime difficulty for anyone building a CE model for health care analysis is uncertainty. The same drug often does not produce the same impact on two apparently similar patients. Factors such as patient age, gender, condition severity, and other health characteristics may impact a drug's performance. Some of these factors are explicitly measurable and accountable; many are not. For example, a 10 mg dose of Drug A does not exactly improve every patient's condition for a specific index by 12 percent. Some patients get measurably (but variably) better, some have adverse drug reactions, and some patients may receive no effect or even get worse.

Because of this uncertainty, a modeler is often required to create the prospective patient population by probabilistically assigning unique patient characteristics, treatment responses, and state changes (e.g., natural death) that may vary as each patient moves through the model of care actions and possible outcomes. The use of probabilistic modeling techniques raises a fundamental analysis question: Are the differences between two alternatives the result of statistical noise (i.e., the probability sampling done by the simulation model) or the true effects of the alternatives being evaluated?

Address correspondence to Daniel R. Murphy, M.S., Medical Decision Modeling Inc., 8909 Purdue Road, Suite 550, Indianapolis, IN 46268; e-mail: drmm@mdm-inc.com. Robert W. Klein, M.S., Lee J. Smolen, B.S., and Timothy M. Klein, B.S., are with Medical Decision Modeling Inc., Indianapolis, IN. Stephen D. Roberts, Ph.D., is with the North Carolina State University, Raleigh, NC.

Isolating Treatment Effects from Statistical Noise

The tools available to a modeler to isolate true effects from statistical noise are called variance reduction techniques (VRTs). The two simplest VRTs are to:

1. Simulate more patients and/or
2. Simulate many replications of the model each with different sample cohorts

How well these techniques work in practice depends on the complexity of the random processes modeled, how those processes are modeled, and the number and types of random processes within the model. A typical approach is to simulate more patients and/or make more repetitions until some critical output measures (e.g., average patient cost) stabilize around a (hopefully) true mean value. Stability is typically measured by the standard error of the mean or through a confidence interval. Of course, the simulation of more patients requires more computing time.

More sophisticated approaches to variance reduction are described in the simulation literature (Bratley, Fox, and Schrage 1987; Law and Kelton 2000). One of those techniques, common random numbers (CRNs), has received recent attention in patient-level simulation from Shechter et al. (2006) and Stout and Goldie (2008). Two fundamental ideas are encompassed in using CRNs: (1) to create identical patient populations across treatment arms and (2) to eliminate the differences between treatment arms that are not due to actual differences in the comparators.

To create common patients, individual members of the simulated heterogeneous cohort must be *exactly* the same across treatment arms. Two populations with the same mean patient age and gender distribution in a patient-level microsimulation may not be identical. In a common patient cohort, for example, the age and gender of the eighth patient must be the same. This correspondence holds true for every patient characteristic of every patient across every treatment arm evaluated. It implies that the same random numbers (RNs) are used for the same characteristics by the same patient across all treatment arms.

A necessary, but not sufficient, condition to insure that the same random numbers/variates are used in different treatment arms is to assign a different random number stream (with a different initial random number seed) to each source of statistical variability in the simulation model. This way each characteristic (e.g., starting age, time on treatment) is taken from the same random number source. Identical random numbers across cohorts may still not guarantee common patients because of the lack of event synchronization. For

example, if a patient dies in an operation, then any random numbers to assign subsequent treatment characteristics are obviously unneeded, but if that operation succeeds (due to other factors) in another cohort, then those subsequent treatment characteristics will be sampled.

To illustrate the use and value of CRNs, consider a hypothetical three-cycle simulation of a single patient comparing treatments D_1 and D_2 . Sampled random numbers RN_1 , RN_2 , etc. are used to determine patient starting ages and cycle outcomes. A patient's cycle outcome is a function of the treatment and the patient's age. Treatment D_2 is more effective than D_1 . The patient either lives or dies during a given cycle. Figure 1 presents model implementations with and without the use of CRNs.

In the top portion of Figure 1 (without CRNs), results are displayed where sequential random numbers are sampled as needed. The patient's starting age for treatment D_1 is sampled using random number RN_1 . The patient lives in Cycles 1 and 2 and dies in Cycle 3. With treatment D_2 , the patient's starting age is determined by RN_2 , and the patient dies in Cycle 2.

Without using CRNs, the patient has a different starting age and the simulation uses different sampled random numbers for cycle outcomes for each treatment. Because of this, the ability to distinguish the differential treatment effects of patients is diminished. A large number of patients would need to be simulated to determine the true difference in treatment outcomes.

In the middle portion of Figure 1, results are presented with the full use of CRNs. The patient begins with the same starting age across treatments D_1 and D_2 , because he or she uses the same random number (RN_1). Both patients use the same sampled random numbers across treatments D_1 and D_2 to determine cycle outcomes. Note that for D_1 using full CRNs, the random number RN_4 is struck through (e.g., RN_4) because the patient died in the previous cycle, which cannot be known a priori. These unused random numbers maintain both the sequence of samples and the event synchronization required for the full implementation of CRNs.

The use of full CRNs eliminates sampling-related variations in patient ages and treatment effects. The age of the patient is the same across treatment arms; thus, for each cycle any age-based differences in treatment effects and/or baseline risks are eliminated. In addition, identical random numbers are used for cycle outcomes for both treatments. Thus, the remaining difference between treatment arms is due entirely to the modeled difference between D_1 and D_2 treatment effects. This reduction in sampling-related variations reduces the number of patient simulations required to determine the relative difference in treatment outcomes (to a given statistical significance).

Figure 1: A Single Patient Example of Treatment Evaluation

A partial implementation of CRNs, creating common initial patient characteristics (patient age in our example), is shown at the bottom of Figure 1. This method of partial synchronization is frequently used and relatively easy to implement. Using partial CRNs, the patient in each treatment arm has the same starting age based on RN_1 . The cycle-by-cycle outcomes, however, are based on the use of sequential random numbers that differ between

treatment arms. In our example, the patient on treatment D_2 lives one cycle less compared to the full CRN example. The impact of partial CRNs on the number of simulated patients required to determine the relative difference in treatment outcomes depends on the relative impact of patient age on cycle-by-cycle outcomes for each treatment.

Our example simulates a simple situation with only one patient and one patient attribute (age) for illustrative purposes. Even without the use of CRNs, simulating an increasing number of patients will eventually result in an acceptable estimate of the relative difference between treatment outcomes. The use of full CRNs, however, will produce acceptable estimates with fewer simulated patients due to reduction in sampling-related variations. The impact of the use of partial CRNs on the required number of simulated patients is indeterminate and simulation dependent.

METHODS

Model Description: Osteoporosis Treatment

A Monte Carlo java-based microsimulation model to evaluate osteoporosis treatments was adapted for use to elucidate the effectiveness of CRNs. The model, using a full implementation of CRNs, was originally created to estimate the long-term clinical and cost-effectiveness of teriparatide use in high-risk osteoporosis patients in Sweden.

The patient cohort simulated consists of male patients with idiopathic male osteoporosis as well as female patients with severe postmenopausal osteoporosis. The simulated patient cohort is defined by these initial characteristics:

1. Starting age (between 60 and 75 years),
2. 88 percent female, 12 percent male,
3. Initial bone mineral density (BMD) T-scores of -3.0 standard deviations or lower than the young adult peak mean, as measured at the femoral neck (proximal femur), and
4. Two previous vertebral fractures, one occurring within 6 months prior to entry into the model.

The model simulates patient interventions with an 18-month regimen of daily teriparatide use as a first-line treatment. It estimates accumulated costs associated with treatment and the costs associated with fracture events. Patient

utilities are applied to estimate quality-adjusted life years (QALYs). For this analysis, the model compares the cost-effectiveness of teriparatide treatment against no treatment. The patient flow associated with the model is shown in Figure 2.

Model inputs are detailed in Table A1 in Appendix A.

Implementation of Common Random Numbers

To implement CRNs, a separate patient cohort was probabilistically created for each replication of the simulation. These identical initial cohorts were maintained by the model and simulated across both the teriparatide and no treatment arms for each replication.

Within the model, each patient had an array of random numbers generated for each process (e.g., probability of a hip fracture) requiring a sample. The number of samples generated was sufficient to ensure that patients received a new random number when needed for a process. A by-product

Figure 2: Osteoporosis Model Patient Flow

of this implementation is that samples for different processes use more random numbers in their random number stream than the model explicitly requires.

Analysis of the Effects of Commonality in Patient Cohorts

In our implementation of CRNs, two options were evaluated: (1) a full implementation of CRNs involving identical patients, common random numbers, and synchronized processing for each patient across the teriparatide and no treatment arms, and (2) a partial implementation of CRNs where only the initial patient characteristics (e.g., age, gender, T-score for a specific patient) are identical across both arms. In the partial implementation of CRNs, the random processes for patients once they entered the simulation are not coordinated between cohorts.

Risk analysis looked at variations in outcomes from 1,000 simulations of 1,000 patients. While initial patient populations changed, the cost and service parameters associated with the underlying model remained at their base case values. This analysis provides a range of expected outcomes that might represent a large hospital or health service area caring for a population of osteoporosis patients. The purpose of such an analysis is to provide estimates on likely best-and worst-case scenarios.

Scatter plots and an acceptability curve developed from 1,000 simulations of 1,000 patients are presented. Unlike a traditional probabilistic sensitivity analysis (PSA), however, the parameter values are not sampled to generate this curve. The acceptability curve we present is based solely on the random numbers sampled against the base case parameter values during each simulation. This approach isolates the sampling variation from the parameter variation, which for most PSAs is controlled by the selection of the variables and the range of variation selected by the modeler.

RESULTS

Base Case Results

Figure 3 shows model results for the Average Patient Cost Difference between patients on teriparatide treatment and patients not treated. Although the analysis was performed from a Swedish health care perspective, and cost inputs were in Euros, for this manuscript the differences in U.S. dollars (\$) are plotted against the number of patients simulated. The figure shows results from single

Figure 3: Average Patient Cost Differences: Teriparatide versus No Treatment

long simulations made without CRNs, using the partial implementation of CRNs, and one with full CRNs implemented.

Looking at Figure 3, two observations are apparent:

1. Full CRNs produced results with significantly less variation around the long-term mean than either of the other strategies.
2. Partial CRNs produced no significant variance reduction over No CRNs. In fact, for this series of simulations, Partial CRNs are the slowest to stabilize around the mean incremental cost. The authors ran additional simulations (with different random number seeds) to investigate these results. Uniformly, Partial CRNs performed only marginally better than No CRNs for this model.

Table 1 compares the statistics for the incremental patient costs of using teriparatide versus no treatment for each strategy for a simulation of 1,024,000 patients. The full use of CRNs reduces the variance of the mean incremental cost of patients using teriparatide versus no treatment by 93.6 percent and reduced the half width of the confidence interval (CI) by 74.7 percent. The impact of the Partial CRNs strategy (i.e., common initial patient characteristics

Table 1: Comparison of Variance Reduction Strategies for Incremental Patient Costs

	Full CRNs	Partial CRNs	No CRNs
Incremental patient costs (\$)			
Mean	1,401	1,394	1,397
SD	23,189	88,956	91,530
Variance reduction	93.58%	5.55%	
95% CI half width reduction	74.67%	2.81%	

CI, confidence interval; CRNs, common random numbers.

Figure 4: Comparison of Ninety-Five Percent Confidence Intervals with and without Common Random Numbers (CRNs)

only) was marginal, showing a variance reduction of 5.6 percent and a half width reduction of 2.8 percent over runs with no variance reduction strategies implemented.

Figure 4 shows the 95 percent CIs for the difference in patient costs of using teriparatide versus no treatment with Full CRNs and No CRNs. The CI for the Partial CRNs strategy was not included because it essentially tracks the CI for No CRNs. These CIs were plotted against the number of patients simulated (on a geometric scale) and the variation from the mean value (i.e., 0 on the vertical axis represents the mean). From the figure, the same CI achieved simulating 1,000 patients using Full CRNs required simulating 16,000 patients without using CRNs.

Risk Analysis Results

Many important decisions are based on risk analysis as well as likely costs and benefits. Another way to look at the value of CRNs is to illustrate how removing extraneous variations provides a clearer picture of such risks. For this model, the Total Incremental Costs and Total QALYs for 1,000 simulations of 1,000 patients were plotted.

Figure 5 shows scatter plots of the distribution of total costs and benefits (1,000 pairs of data) with and without CRNs. Both scatter plots show each pair of Total Incremental Costs and QALYs for each replication and the Mean value point for the 1,000 replications. The plot for No CRN's displays a significantly more risky appraisal of possible outcomes across both measures than the plot using Full CRN's. For this model, Figure 5 suggests that the teriparatide usage in this high-risk population is more effective than no treatment.

From Figure 5, 660 simulations produced positive costs and QALYs without CRNs, compared with 975 simulations using CRNs. Using CRNs, all 1,000 simulations fall into quadrants with positive incremental QALYs when comparing teriparatide use versus no treatment. With No CRNs, 88 of the 1,000 simulation replications estimated negative incremental QALYs. In terms of patient outcomes, the full use of CRNs (in this model) estimates a more favorable set of impacts for the use of teriparatide, particularly if the decision maker is concerned with maximum regret (Smith 1996) (i.e., the consequences of the worst-case scenario). The mean Cost per QALY did not change, but the estimated probability of a better outcome did improve. The use of CRNs better isolated the impact of the drug.

Figure 5: Risk Analysis: Scatter Plots of Total Increment Costs and Quality-Adjusted Life Years (QALYs) with and without Common Random Numbers (CRNs)

Impact of CRNs on Willingness to Pay

Acceptability curves are frequently used to display and characterize net benefits or a cost per effectiveness measure from multiple replications of a simulation as a function of willingness to pay (WTP). For this model, the Total Incremental Costs per QALY gained for teriparatide use versus no treatment were used. To generate the acceptability curves, data were collected from 1,000 simulations of 1,000 patients using No CRNs, Partial CRNs, and Full CRNs. The acceptability curve is determined by the cumulative percentage of simulations whose Cost per QALY is less than a given value for a given WTP.

The threshold line representing a given WTP cuts through two quadrants (positive QALYs and positive Costs; negative QALYs and negative Costs). As the WTP changes, the percentage of simulations less than the specified WTP will change. This change may not be monotonic as the threshold value increases; it depends on the location of the points in both the quadrants being cut by the WTP line (consider a WTP line rotating through these two quadrants in Figure 5, for example). As the Cost is less than zero, the simulation replications that produced positive QALYs and negative Costs will be less than any WTP (dominant). Those replications representing negative QALYs and positive Costs will always be greater than any WTP (dominated).

The acceptability curves for a selected set of threshold values using Full CRNs and No CRNs are shown in Figure 6. The acceptability curve for Partial CRNs is not shown as it tracks within 3 percent of the No CRNs curve. Using Full CRNs, the 75.1 percent level is reached at a WTP of \$10,000 per QALY. Moreover, because 78 of the 1,000 samples had negative QALYs, only 92.2 percent would be acceptable with an unlimited willingness to pay when CRNs are not used.

DISCUSSION

Along with the value of CRNs, a few specific points merit some additional discussion. The relatively low variance reduction reported for this model for Partial CRNs is not reflective of the authors' previous work nor that reported from other studies (Shechter et al. 2006; Smolen, Klein, and Klein 2010). The explanation for the relative lack of variance reduction is a function of three features of the model. First, the number and intervals of the cycles used within the model allow any differences related solely to initial patient characteristics to be absorbed (in effect) by the model quickly. Second, the

Figure 6: Acceptability Curves with and without Common Random Numbers (CRNs)

initial cohort is relatively homogeneous. A cohort with a larger age range and higher maximum initial bone mineral density (BMD) T-score, for example, would produce more variance from the initial conditions. In that case, the expectation for a more significant reduction in variance using Partial CRNs (i.e., common initial conditions) would be greater. Finally, in this model, death is evaluated as a cycle-by-cycle probability as opposed to having a time of death (or remaining natural lifetime) sampled as an initial patient characteristic. Total life years are usually the biggest driver of QALYs. Had we modeled (sampled) each patient's time of death as an initial condition, the variance reduction for Partial CRNs would be closer to that for Full CRNs.

As noted by Stout and Goldie (2008), there may be specific variables for which CRNs between treatments are inappropriate. This can occur, for example, when the variable of interest is treatment-related. If the treatments act by different mechanisms, then there is no reason to believe that relative efficacies of the treatments or propensities for treatment-related side effects on a particular patient are maximally correlated. In such cases,

a separate series of random numbers should be applied for each treatment.

In addition, CRNs may be inappropriate where model interactions of multiple variables exist. In these cases, the use of CRNs can actually increase variance (Wright and Ramsay 1979). The value of CRNs is model specific: it can only be determined (except in simple cases) by comparing model variance estimates with and without the use of CRNs. Our general recommendation for anyone using CRNs is to execute these test runs with their model.

Figure 3 illustrates that arbitrarily selecting a number of patients for a complex simulation like this model can produce misleading results. Until the variability in the costs and WTP estimates stabilize, the ability to make complete probabilistic statements about the cost-effectiveness of treatment is questionable. Finding that point of stability (e.g., the number of patients and/or the length of a simulation) requires the modeler to (1) examine incremental results like those displayed in Figure 3 or (2) use an adaptive algorithm to monitor and determine when statistical stability is achieved.

In Figure 5, the parameter values used for the base case model were unchanged; only the sample batch size (1,000 batches of 1,000 patients) changed from the base case in the risk analysis. The idea behind this application of probabilistic sensitivity analysis was to isolate the impact of sampling variation from (combined) parameter and sampling variation, which is the method to implement a typical PSA. The shape and variation in a typical PSA is determined by the specific variable selected for analysis and the ranges and types of variations selected for each individual simulation run. The differences in the two plots shown in Figure 5 are solely a function of the use of CRNs.

The acceptability curves in Figure 6 represent a measure of how sampling variability can impact willingness to pay for groups of the number of patients simulated. They would not be reflective of a hospital that sees 20 such patients per year nor a country's health plan that covers 50,000 patients.

The acceptability curves presented also illustrate the policy impact of using CRNs. First, the acceptability curve for the treatment is steeper with CRNs. This indicates that the willingness to pay has less variation than without CRNs (a WTP curve that climbs at a slower rate). These differences reflect on the relative statistical confidence a decision maker can place on the estimates generated by the model.

Health care analysis is becoming more complex and more computationally intensive. The increasing interest in the estimation of the expected value of perfect, partial perfect, and sample information (EVPI, EVPPI, and EVSI)

is a case in point (Barton, Briggs, and Fenwick 2008). These techniques require a series of simulations, typically involving 1,000 or more replications of thousands of patients. The use of CRNs can substantially reduce the computational load for EVPI-type analyses and make the use of these techniques more accessible.

CONCLUSIONS

Simulation models used in medical decision analysis contain many elements that suggest the use of CRNs. These models typically involve two or more patient treatment arms, large patient cohorts, multiple random processes, probabilistically assigned patient characteristics, and frequent periodic probabilistic state changes. CRNs are especially effective in reducing the number of simulated patients required to achieve a given level of statistical confidence.

A partial implementation of CRNs, used to generate identical initial patient cohorts as done by default in software such as TreeAge, was only marginally effective in reducing variance in this analysis. We believe this result is attributable to the specific structure of the model we created. If key drivers of effectiveness, such as patient natural lifetime, were sampled up front, we believe that partial CRNs would produce greater variance reductions, as shown in other studies.

Our work indicates that benefits of CRNs extend beyond improvements to simulation efficiency. The use of CRN's can provide information on the discrete distribution of the number of patients that benefit or are harmed by a treatment as well as its mean benefit. Using language from a recent taxonomy of uncertainty, for any particular number of simulated patients using CRNs in a model reduces the ambiguity surrounding the cost-effectiveness estimate (Han, Klein, and Arora 2011). In effect, risk estimates (attributable to treatment) are overestimated without CRNs, because identical treatments do not produce identical results when the cohorts differ. For a decision maker, this implies a more concentrated range of possible outcomes than those not using CRNs.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The model development and analysis work reported in this article was funded by Eli Lilly & Company (Lilly), Indianapolis, Indiana, USA. Medical Decision Modeling (MDM) does not

have any contractual obligation to submit this manuscript to Lilly for pre-publication approval. Lilly has reviewed this manuscript solely to ensure that no confidential information was disclosed. The model, analyses, and conclusions presented in this manuscript were developed solely by the authors.

MDM is a contract research organization that performs pharmaco-economic research to evaluate current and prospective treatments for medical conditions. Lee J. Smolen and Robert W. Klein are equity partners of MDM.

The authors wish to acknowledge Russel Burge, Ph.D., Global Health Outcomes, Eli Lilly Lilly & Company, for his assistance in the preparation and review of this manuscript.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Barton, G. R., A. H. Briggs, and E. A. L. Fenwick. 2008. "Optimal Cost-Effectiveness Decisions: The Role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfection Information (EVPI)." *Value Health* 11 (5): 886–97. doi: 10.1111/j.1524-4733.2008.00358.x
- Barton, P., S. Bryan, and S. Robinson. 2004. "Modelling in the Economic Evaluation of Health Care: Selecting the Appropriate Approach." *Journal of Health Services Research and Policy* 9 (2): 110–8. doi: 10.1258/135581904322987535
- Boyd, K. A., E. Fenwick, and A. Briggs. 2010. "Using an Iterative Approach to Economic Evaluation in the Drug Development Process." *Drug Development Research* 71 (8): 470–7. doi: 10.1002/ddr.20421
- Bratley, P., B. L. Fox, and L. E. Schrage. 1987. *A Guide to Simulation*. Berlin: Springer-Verlag.
- Brennan, A., S. E. Chick, and R. Davies. 2006. "A Taxonomy of Model Structures for Economic Evaluation of Health Technologies." *Health Economics* 15 (12): 1295–310. doi: 10.1002/hec.1148
- Burman, C. F., and S. J. Wiklund. 2011. "Modelling and Simulation in the Pharmaceutical Industry—Some Reflections." *Pharmaceutical Statistics* 10 (6): 508–16. doi: 10.1002/pst.523
- Caro, J. J., A. H. Briggs, U. Siebert, and K. M. Kuntz. 2012. "Modeling Good Research Practices – Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1." *Medical Decision Making* 32 (5): 667–77.
- Dental and Pharmaceutical Benefits Agency. n.d. "TLV Historical Drug Prices." *Dental and Pharmaceutical Benefits Agency* [accessed on March 8, 2012]. Available at <http://www.tlv.se/beslut/sok/lakemedel/>

- Genant, H. K., J. Halse, W. G. Briney, L. Xie, E. V. Glass, and J. H. Krege. 2005. "The Effects of Teriparatide on the Incidence of Back Pain in Postmenopausal Women with Osteoporosis." *Current Medical Research and Opinion* 21 (7): 1027–34. doi: 10.1185/030079905X49671
- Han, P. K. L., W. M. P. Klein, and N. K. Arora. 2011. "Varieties of Uncertainty in Health Care: A Conceptual Taxonomy." *Medical Decision Making* 31 (6): 828–38. doi: 10.1177/0272989X11393976
- Johnell, O., J. A. Kanis, A. Oden, I. Sernbo, I. Redlund-Johnell, C. Pettersson, C. De Laet, and B. Jonsson. 2004. "Mortality after Osteoporotic Fractures." *Osteoporosis International* 15: 38–42.
- Kanis, J. A., O. Johnell, A. Oden, I. Sembo, I. Redlund-Johnell, A. Dawson, C. De Laet, and B. Jonsson. 2000. "Long-Term Risk of Osteoporotic Fracture in Malmö." *Osteoporosis International* 11 (8): 669–74.
- Kanis, J. A., M. Stevenson, E. V. McCloskey, S. Davis, and M. Lloyd-Jones. 2007. "Glucocorticoid-induced Osteoporosis: A Systematic Review and Cost-utility Analysis." *Health Technology Assessment* 11 (7): iii–iv, ix–xi, 1–231.
- Law, A. M., and W. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd Edition. New York: McGraw-Hill.
- Lindsay, R., W. H. Scheele, R. Neer, G. Pohl, S. Adami, C. Mautalen, J. Reginster, J. J. Stepan, S. L. Myers, and B. H. Mitlak. 2004. "Sustained Vertebral Fracture Risk Reduction after Withdrawal of Teriparatide in Postmenopausal Women with Osteoporosis." *Archives of Internal Medicine* 164 (18): 2024–30. doi: 10.1001/archinte.164.18.2024
- Lundberg, L., M. Johannesson, D. G. L. Isacson, and L. Borgquist. 1999. "Health-State Utilities in a General Population in Relation to Age, Gender and Socioeconomic Factors." *The European Journal of Public Health* 9 (3): 211–7. doi: 10.1093/eurpub/9.3.211
- Lundkvist, J., O. Johnell, C. Cooper, and D. Sykes. 2006. "Economic Evaluation of Parathyroid Hormone (PTH) in the Treatment of Osteoporosis in Postmenopausal Women." *Osteoporosis International* 17 (2): 201–11. doi: 10.1007/s00198-005-1959-4
- Neer, R. M., C. D. Arnaud, J. R. Zanchetta, R. Prince, G. A. Gaich, J. Y. Reginster, A. B. Hodsman, E. F. Eriksen, S. Ish-Shalom, H. K. Genant, O. Wang, and B. H. Mitlak. 2001. "Effect of Parathyroid Hormone (1–34) on Fractures and Bone Mineral Density in Postmenopausal Women with Osteoporosis." *The New England Journal of Medicine* 344 (19): 1434–41. doi: 10.1056/NEJM200105103441904
- NICE. 2008. "The Guidelines Manual 2009 – Chapter 7: Assessing Cost Effectiveness" December 24 [accessed on July 25, 2012]. Available at <http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/GuidelinesManual2009.jsp?domedia=1&mid=68D113F7-19B9-E0B5-D4CA25CD07AAA3E6>
- Petrou, S., and A. Gray. 2011. "Economic Evaluation alongside Randomised Controlled Trials: Design, Conduct, Analysis, and Reporting." *British Medical Journal* 342: 1–6. doi: 10.1136/bmj.d1548

- Prince, R., A. Sipos, A. Hossain, U. Syversen, S. Ish-Shalom, E. Marcinowska, J. Halse, R. Lindsay, G. P. Dalsky, and B. H. Mitlak. 2005. "Sustained Nonvertebral Fragility Fracture Risk Reduction after Discontinuation of Teriparatide Treatment." *Journal of Bone and Mineral Research* 20 (9): 1507–13. doi: 10.1359/JBMR.050501
- Shechter, S. M., A. J. Schaefer, R. S. Braithwaite, and M. S. Roberts. 2006. "Increasing the Efficiency of Monte Carlo Cohort Simulations with Variance Reduction Techniques." *Medical Decision Making* 26 (5): 550–3. doi: 10.1177/0272989X06290489
- Smith, R. D. 1996. "Is Regret Theory an Alternative Basis for Estimating the Value of Healthcare Interventions?" *Health Policy* 37 (2): 105–15.
- Smolen, L. J., R. W. Klein, and T. M. Klein. 2010. "Synchronization of Random Number Streams Greatly Enhances Efficiency of Probabilistic Models." *Value in Health* 13 (3): A16.
- Statistics Sweden. n.d. "Sweden Life Tables for 2003–2007" [accessed on December 2, 2011]. Available at <http://www.scb.se/>
- Stout, N. K. and S. J. Goldie. 2008. "Keeping the Noise Down: Common Random Numbers for Disease Simulation Modeling." *Health Care Management Science* 11 (4): 399–406.
- Wright, R. D., and T. E. Ramsay. 1979. "On the Effectiveness of Common Random Numbers." *Management Science* 25 (7): 649–56. doi: 10.1287/mnsc.25.7.649

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Table A1: Model Input Data (Abbreviations: Rx, Prescription; FX, Fracture).