LEADING ARTICLE

Dis Manage Health Outcomes 2007; 15 (1): 7-12
1173-8790/07/0001-0007/$44.95/0

Estimating the Effect of Regression to the Mean in Health Management Programs

Ariel Linden[1,2]

1 Linden Consulting Group, Portland, Oregon, USA
2 School of Medicine and School of Nursing, Oregon Health & Science University, Portland, Oregon, USA

Abstract

Most health management programs, such as disease management or health promotion/wellness interventions, implement targeted interventions for an identified high-risk group, leaving the remaining non-managed lower-risk population as controls. This is problematic from an outcomes perspective because individuals initially identified by their high-risk scores will inevitably have lower average scores on remeasurement, even in the absence of a health management program. This statistical phenomenon is called regression to the mean (RTM). This article presents actual examples of RTM, describes the classic method for estimating the impact of RTM in a pre-post study, and provides suggestions for designing health management program evaluations to mitigate the effects of RTM.

Most health management programs, such as disease management (DM) or health promotion/wellness interventions, are implemented at the population level. Generally, a target population is first determined and then individuals at highest risk within that population are identified. Those persons are then invited to enroll in an intensive intervention aimed at improving their health. In DM, 'high risk' typically refers to individuals who have accrued high medical costs in the past year, whereas wellness programs may classify 'risk' based on health survey results or current clinical indicators (e.g. blood pressure, laboratory test values).

Given this type of program design, the ability to employ controlled evaluation studies is limited. Most individuals identified as high risk enroll in the program intervention at some point, leaving the remaining non-managed lower-risk population as controls. This is problematic from an outcomes perspective because a group initially identified by their high-risk scores will inevitably have lower average scores on remeasurement, even in the absence of a health management program. This statistical phenomenon, originally termed 'regression toward mediocrity', was described over a century ago by Sir Francis Galton upon discovering that on average, tall parents had children shorter themselves and short parents had taller children.[1] An excellent historical review of regression to the mean (RTM) is provided by Stigler.[2]

RTM is the result of random fluctuation or non-systematic error in repeated measurement. A simple example of this occurs in measuring blood pressure or heart rate. Rarely are any two observations identical, even if taken minutes apart. At the individual level this is called within-subject variability. Within this construct, higher (or lower) initial values are likely to be followed by an observation closer to the person's average value or score. Over the course of many repeated observations, this variability narrows around the true mean.[3,4] Grouping individuals according to their initial outlier measurement exacerbates the effect of RTM. Similar to individual level measures, groups with high (or low) initial mean values will tend to regress to the mean of the overall group. Thus, in the evaluation of a pre-post study, the researcher must be able to separate the RTM effect from the true treatment effect in order to accurately determine the effectiveness of the intervention.

The implications of RTM on the measurement of specific health-related outcomes has been examined extensively in the literature and has substantiated the need to address RTM.[5-16] However, this concern is rarely addressed in pre-post program evaluations of health management programs. Consequentially, RTM remains a major source of bias pervasive in the evaluation designs of the industry.[17]

This article begins by presenting real examples of RTM to illustrate the phenomenon. Next, the classic method for estimating the impact of RTM in a pre-post study will be described so that this technique can be easily replicated in health management programs. Finally, suggestions will be presented for designing health management program evaluations to mitigate, or at least account for, the effects of RTM.

1. Real Examples of Regression to the Mean (RTM)

RTM can easily be mistaken for a program effect in the absence of an equivalent control group. The best approaches to illustrate the effect of RTM are either by using observations taken from time periods in which no health management programs were implemented or by using control-group data derived from a research study. In this section, both situations will be presented.

Figure 1 presents the average annual costs (adjusted to 2005 values) of all health plan members with coronary artery disease (CAD) who were continuously enrolled for 4 years (2001–4) in a medium-sized California health plan. Over this period no health management programs were introduced. The high-risk group is comprised of members originally identified as being in the top quintile in costs for 2001 and the low-risk group includes everyone else during that initial measurement year. As shown, mean costs for that high-risk group dropped precipitously from 2001 to 2002 (by approximately $US24 000), while mean costs of the remaining members during that same period rose by only $US920. This scenario perfectly illustrates the effect of RTM. Had a DM program targeting CAD existed during this period, an evaluation of the impact on costs would have wrongly concluded that this outcome was a program effect.

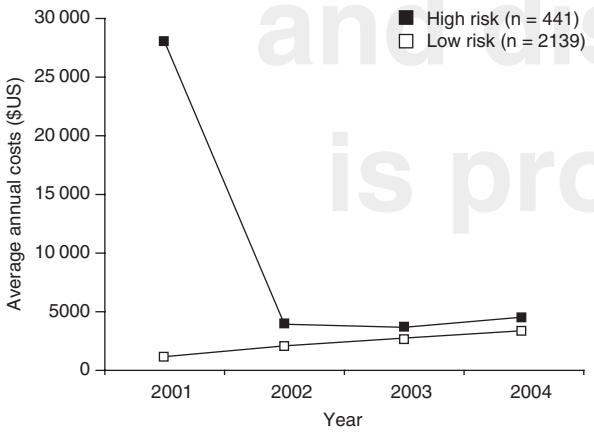

Fig. 1. Annual costs of coronary artery disease patients who were enrolled in a health plan for 4 continuous years. The high-risk group is comprised of the top quintile in costs for 2001 and the low-risk group includes everyone else during that initial measurement year. All costs are adjusted to 2005 values.

Figure 2 presents physical component summary (PCS) scores from the short form-12 (SF-12) health status survey[18] for a control group from a study conducted at a large organization in the Northwest of the US.[19] Scale values are standardized from 0 to 100, with higher values indicating better physical health. For illustrative purposes, the high-risk group is classified as having a PCS score of <44.25, which corresponds to the 25th percentile at the US national level.[20] Control group participants were surveyed twice: once at program commencement and then again at 3 months. They received no intervention. As shown, the high-risk group increased their PCS score by >8 points (22.6%), while the lower-risk group remained unchanged. Once again these findings clearly illustrate an RTM effect, as the high-risk group had a mean value on remeasurement that was closer to the overall mean.

2. Estimating the Effect of RTM

As demonstrated in the previous section, RTM poses a serious threat to the validity of any pre-post evaluation. While a number of statistical models have been produced to account for the magnitude of the RTM effect,[3,21-34] the most widely used model was developed more than 30 years ago.[3,21] The model uses an iterative, which begins by calculating the z-score (equation 1):

$$z = (\kappa - \mu)/\sigma \quad \text{(Eq. 1)}$$

where μ is the baseline population mean, σ is the standard deviation (SD) of the entire sample, and κ is the cutoff score representing the high-risk or target group value. Next, the c-statistic is calculated (equation 2):

$$c = \varphi(z)/([1 - \phi(z)]) \quad \text{(Eq. 2)}$$

where $\varphi(z)$ is the probability density function for z, and $\phi(z)$ is the cumulative distribution function for z. The expected first measurement value for the target group is calculated as (equation 3):

$$\text{Expected first mean value} = \mu + (c\sigma) \quad \text{(Eq. 3)}$$

The expected value of the follow-up measurement is (equation 4):

$$\text{Expected second mean value} = \mu + (c\sigma\rho) \quad \text{(Eq. 4)}$$

where ρ is the Pearson within-subject correlation between all values of both periods. Thus, the expected RTM effect is given by subtracting the expected first measurement value from the second

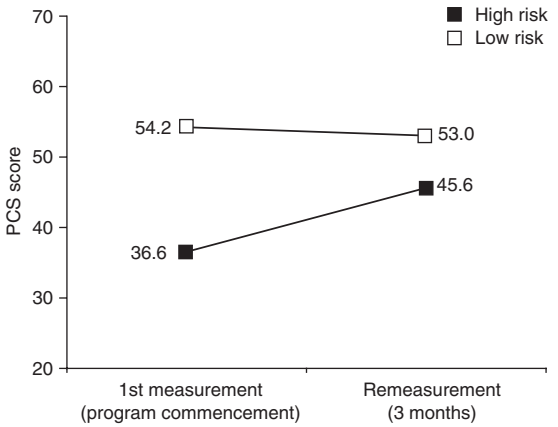

Fig. 2. Physical component summary (PCS) scores on the short form-12 from a control group participating in a health coaching study.[19] Scale values are standardized from 0 to 100, with higher values indicating better physical health. All participants were surveyed twice: once at program commencement and then again at 3 months. Squares represent mean scores.

(or vice versa, depending on the expected direction of the variable).

Given that this is a statistical model, program administrators and evaluators should be cognizant of factors that can influence the results of these calculations. For example, the RTM effect increases as a function of increased variability in the measure under study. Additionally, the further the cutoff point is from the mean and the weaker the correlation is between pre-post measurements, the greater the RTM effect is. More often than not, both factors will move in tandem, so that setting the cutoff score closer to the mean will increase the correlation between measurement periods. However, it may not be reasonable to change the cutoff point as it may be set to a level intended to identify a certain percentage of the population at risk.

Another concern is with the distribution of the outcome variable under study. This model was developed for normally distributed data, and, therefore, may produce spurious results with non-normal data. While other models have been developed to deal with non-normal distributions,[28,32] transforming the data (i.e. logarithms) before using the model presented here may suffice.

3. Applying the RTM Model to Health Management

To demonstrate the applicability of this model to a health management program, we will revisit the 2001–2 data used in figure 1. A test for normality[35] indicated that the cost data required transformation to meet the criteria for inclusion in the RTM estimator; therefore, the natural log of each observation was taken. Next, the mean of the transformed values (7.61), SD of the transformed values (1.89), and correlation between first and second year transformed observations (0.27) were calculated. The cutoff value was set as the first observation of the transformed fifth quintile = 9.40. Thus, z = 0.95, c = 1.48, the expected 2001 high-risk group transformed mean = 10.4 (untransformed = $US33 199), and the expected 2002 high-risk group transformed mean = 8.4 (untransformed = $US4245). As shown in figure 3, the 2002 expected high-risk group untransformed mean value is similar to the actual 2002 observation ($US3932), while the expected 2001 untransformed mean value is higher than the actual observation ($US28 000). This difference is most likely explained by the few outlier costs that even log transformation could not smooth out. Nonetheless, the model appears to predict the estimated RTM effect in the high-risk category rather well.

Some health management programs contend that even though they do not directly intervene on the low-risk group, the costs of this group are impacted by influencing physician behavior vis-à-vis a 'spill-over' effect from the high-risk group (e.g. complying with guidelines). Using the RTM model we can easily estimate this effect in the low-risk group by modifying the z-score so that the cutoff value is subtracted from the mean (as compared with the high-risk group where the mean is subtracted from the cutoff value). Thus, z = −0.95, c = 0.31, the expected 2001 low-risk group mean = $US1127, and the expected 2002 low-risk group mean = $US1726. As shown in figure 3, both these estimated values are similar to their actual observations, indicating that this model works well to predict the RTM effect for the low-risk group.

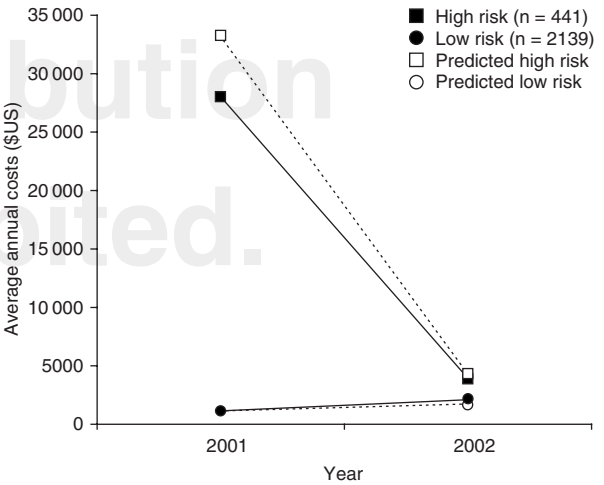

Fig. 3. Comparison of actual and estimated regression to the mean effects for high- and low-risk groups for coronary artery disease who were enrolled in a health plan for 4 continuous years. The high-risk group is comprised of the top quintile in costs for 2001 and the low-risk group comprises the remaining population. All costs are adjusted to 2005 values.

As figure 3 illustrates, the entire population of continuously enrolled individuals is accounted for once the RTM effect is calculated in both high- and low-risk groups. Therefore, to show a program effect adjusting for RTM, a health management program must reduce costs in the high-risk group beyond that estimated by the model, while ensuring that costs do not increase in the low-risk group beyond the model estimates.

A model to estimate the impact of RTM was described in this section. However, as indicated by the model parameters (in particular, the use of the pre-post correlation coefficient), this estimation is done retrospectively once all the observations are completed and the associated statistical measures have been calculated. Section 4 provides direction in developing strategies prior to program commencement in order to reduce or prevent the impact of RTM on outcomes.

4. Designing Programs to Mitigate the RTM Effect

The most obvious choice in prospective study designs to control for the RTM effect is the randomized, controlled trial (RCT). Random assignment to treatment or control is meant to eliminate selection bias by distributing all known and unknown sources of variation equally between groups. By adding the equivalent control group, the effect of RTM will be accounted for in the difference in outcomes between the groups (i.e. both groups may show an improvement in the health outcome [as a result of RTM] with the differential being the treatment effect). In the health management industry, the use of the RCT is usually limited to research endeavors where strict control over the study environment is possible. However, for the majority of programs, the RCT is neither practical nor feasible in a business setting.

The regression-discontinuity design may be the most suitable study design for health management programs in that it controls for major sources of bias with few design requirements.[36] Individuals are assigned to treatment or control based on a cutoff score of a pre-test variable. Those subjects scoring below the cutoff (assuming a lower score equates to poorer health) are assigned to the intervention while those scoring above the cutoff act as controls. The post-test measure may be the same as or different to the pre-test. In general, the strict adherence to group assignment allows for the determination of an unbiased treatment effect associated with the cutoff value. In other words, individuals closest to the cutoff on either side are similar enough on the outcome measure that any difference between them (discontinuity of the regression line coinciding with the cutoff) would be considered a treatment effect. This can be inspected visually and supported by statistical testing using the analysis of co-variance model (ANCOVA).[32] Others have also suggested the use of ANCOVA to control for RTM;[37] however, without the addition of prospective assignment to treatment or control based on a strict cutoff, the ANCOVA results may still be biased.

As illustrated in figure 1, the effect of RTM is eliminated when several repeated measurements are taken. Therefore, one simple way of controlling for RTM in a prospective manner is to use the average of two or more pre-test measurements as a way of determining patient assignment to the intervention. An enhancement to the model described above can be used, where all parameters are identical except the z-score, which is now changed as follows (equation 5):[3,21]

$$\text{z-score} = [(\kappa-\mu)]/[(\delta^2 + \gamma^2)/n] \quad \text{(Eq. 5)}$$

where κ is the cutoff score, μ is the mean of the multiple observations (n), δ^2 is the variance of true measurement ($p * \sigma^2$), γ^2 is the variance of repeated individual measurements ($\sigma^2 - \delta^2$), and, thus, $[(\delta^2 + \gamma^2)/n]$ is the pooled variance.

Using the CAD cost data, figure 4 illustrates the estimated RTM effect for 2004 using the average of the 2001–3 costs. The mean = $US8339, SD = $US9654, correlation = 0.24, and the cutoff value was set as the first observation of the fifth quintile = $US14 042. As shown, the baseline means of the high- and low-risk groups are closer together than in the previous example using only a 1-year baseline. It should be noted that little benefit is gained by using more than four baseline measurements to estimate the RTM effect.[21]

An important factor to include in the discussion of RTM is the participants' tenure both in the population from which they were

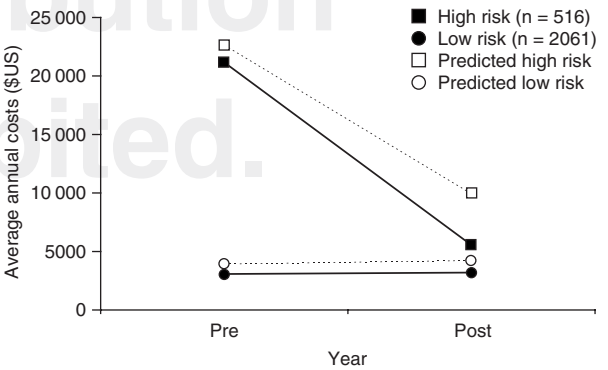

Fig. 4. Comparison of actual and estimated regression to the mean effects for high- and low-risk groups for coronary artery disease using the average of 3 baseline years as the 'pre' value and 1 year as the 'post' value. The high-risk group is comprised of the top quintile in costs for 2001–3 and the low-risk group comprises the remaining population. All costs are adjusted to 2005 values.

drawn as well as in the health management program itself. In health promotion/wellness programs where outcomes (i.e. change in health status or risk) are typically determined via health-risk surveys, the disenrollment of a participant before the post-test measurement results in a loss of data required for the evaluation. There is some evidence to suggest that participants in educational programs who remain enrolled long enough to be re-measured may have poorer markers of health than those not returning.[38] A similar problem exists for DM programs that use population-based outcomes (e.g. utilization metrics and cost). The literature shows that health-plan enrollees with chronic illnesses are much more likely to stay enrolled in the plan than those with no chronic conditions.[39,40] Both cases suggest that participants with initial high-outlier values are more likely to remain for remeasurement than those with lower values, providing further support for the use of either controlled studies or estimation of the RTM effect.

5. Conclusion

This paper illustrates that outcomes from health management programs are particularly vulnerable to the effects of RTM given that (a) individuals are chosen to participate in the intervention based on their outlier baseline 'risk' score, and (b) equivalent controls groups are rarely available for comparison. Under these conditions it is extremely important to isolate the effect of RTM, in order to prevent it from being mistaken for a treatment effect. While it is always preferable to utilize an equivalent-control-group study design to control for bias, this article describes an approach to quantify the RTM effect and demonstrates that the RTM estimator can be a valuable tool when a true control group is unavailable.

Acknowledgments

No sources of funding were used to assist in the preparation of this review. The author has no conflicts of interest that are directly relevant to the content of this review.

References

1. Galton F. Regression towards mediocrity in hereditary stature. J Anthropol Inst 1886; 15: 246-63
2. Stigler SM. Regression towards the mean, historically considered. Stat Methods Med Res 1997; 6 (2): 103-14
3. Gardner MJ, Hardy JA. Some effects of within person variability in epidemiological studies. J Chronic Dis 1973; 26: 781-95
4. Johnson WD, George VT. Effect of regression to the mean in the presence of within-subject variability. Stat Med 1991; 10: 1295-302
5. Boissel JP, Duperat B, Leizorowicz A. The phenomenon of regression to the mean and clinical investigation of blood cholesterol lowering drugs. Eur J Clin Pharmacol 1980; 17: 227-30
6. Andrews G, Harvey R. Regression to the mean in pretreatment measures of stuttering. J Speech Hear Disord 1981; 46: 204-7
7. Shepard DS, Finison LJ. Blood pressure reductions: correcting for regression to the mean. Prev Med 1983; 12: 304-17
8. Denke MA, Frantz ID. Responses to a cholesterol-lowering diet: efficacy is greater in hypercholesterolemic subjects even after adjustment for regression to the mean. Am J Med 1993; 94: 626-31
9. Herpin D, Demange J. Effect of regression to the mean in serial echocardiographic measurements of left ventricular mass: quantification and clinical implications. Am J Hypertens 1994; 7: 824-8
10. Forrow L, Calkins DR, Allshouse K, et al. Evaluating cholesterol screening: the importance of controlling for regression to the mean. Arch Intern Med 1995; 155: 2177-84
11. Prescott RJ, Garraway WM. Regression to the mean occurs in measuring peak urinary flow. Br J Urol 1995; 76: 611-3
12. Pitts SR, Adams RP. Emergency department hypertension and regression to the mean. Ann Emerg Med 1998; 31: 214-8
13. Murakawa Y, Yamashita T, Ajiki K, et al. Ostensible day-night difference of QT prolongation during long-term treatment with antiarrhythmic drugs: reappraisal of the law of 'regression to the mean'. J Cardiovasc Pharmacol 1998; 32: 62-5
14. Cummings SR, Palermo L, Browner W, et al. Monitoring osteoporosis therapy with bone densitometry: misleading changes and regression to the mean. JAMA 2000; 283: 1318-21
15. Asmar R, Safar M, Oueneau P. Evaluation of the placebo effect and reproducibility of blood pressure measurement in hypertension. Am J Hypertens 2001; 14 (6 Pt 1): 546-52
16. Takashima Y, Sumiya Y, Kokaze A, et al. Magnitude of the regression to the mean within one-year intra-individual changes in serum lipid levels among Japanese male workers. J Epidemiol 2001; 11: 61-9
17. Linden A, Adams J, Roberts N. An assessment of the total population approach for evaluating disease management program effectiveness. Dis Manage 2003; 6 (2): 93-102
18. Ware Jr JE, Kosinski M, et al. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996; 34: 220-33
19. Butterworth S, Linden A, McClay M, et al. The effect of motivational interviewing-based health coaching on employees' physical and mental health status. J Occup Healt Psych 2006; 11 (4): 358-65
20. Ware JE, Losinski M, Turner-Bowker DM, et al. How to score Version 2 of the SF-12® health survey (with a Suppl. documenting Version 1). Lincoln (RI): QualityMetric Incorporated, 2002
21. Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. Am J Epidemiol 1976; 104: 493-8
22. Blomqvist N, Cederblad G, Korsan-Bengtsen K, et al. Application of a method for correcting an observed regression between change and initial value for the bias caused by random errors in the initial value. Clin Chem 1977; 23 (10): 1845-8
23. Senn SJ, Brown RA. Estimating treatment effects in clinical trials subject to regression to the mean. Biometrics 1985; 41: 555-60
24. Curnow RN. Correcting for regression in assessing the response to treatment in a selected population. Stat Med 1987; 6: 113-7
25. Senn S. Regression to the mean and crossover trials revisited. Stat Med 1994; 13: 1181-6
26. Hannan PJ, Jacobs Jr DR, McGovern P, et al. Estimating the effect of regression toward the mean under stochastic censoring. Am J Epidemiol 1994; 139 (4): 422-31
27. Ross DC. Using severity rather than randomization for assignment in clinical trials: Monte Carlo tests and practical illustrations of the Robbins-Zhang method. J Psychiatr Res 1995; 29: 315-32
28. Chuang-Stein C, Tong DM. The impact and the implication of regression to the mean on the design and analysis of medical investigations. Stat Methods Med Res 1997; 6: 115-28
29. Chesher A. Non-normal variation and regression to the mean. Stat Methods Med Res 1997; 6: 147-66
30. Copas JB. Using regression models for prediction: shrinkage and regression to the mean. Stat Methods Med Res 1997; 6: 167-83

31. Chen S, Cox C, Cui L. A more flexible regression-to-the-mean model with possible stratification. Biometrics 1998; 54: 939-47
32. Trochim WMK. The research methods knowledge base. 2nd ed. Cincinnati (OH): Atomic Dog Publishing, 2001
33. Muller HG, Abramson I, Azari R. Nonparametric regression to the mean. Proc Nat Acad Sci 2003; 100 (17): 9715-20
34. Tu YK, Baelum V, Gilthorpe MS. The relationship between baseline value and its change: problems in categorization and the proposal of a new method. Eur J Oral Sci 2005; 113 (4): 279-88
35. Rosenthal R. An application of the Kolmogorov-Smirnov test for normality with estimated mean and variance. Psychol Rep 1968; 22: 570
36. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: an introduction to the regression-discontinuity design. J Eval Clin Pract 2006; 12 (2): 124-31
37. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005; 34: 215-20
38. Klepp K-I, Ulvik RJ, Matthiesen SB, et al. Cholesterol screening: observed cholesterol reduction in a prospective study in Norway. Eur J Publ Hlth 1993; 3: 38-42
39. Marquis MS, Buntin MB, Escarce JJ, et al. Is the individual market more than a bridge market? An analysis of disenrollment decisions. Inquiry 2005/2006 Winter; 42 (4): 381-96
40. California HealthCare Foundation. 2005 snapshot: individual health insurance market [online]. Available from URL: http: //www.chcf.org/documents/insurance/SnapshotIndividualMarket05.pdf [Accessed 2006 Nov 24]

About the Author: Dr Linden is a health services researcher whose focus is on evaluating the effectiveness of disease management and health promotion initiatives. In this area alone, he has recently published more than 40 peer-reviewed articles, book chapters, editorials, and a position paper for the Disease Management Association of America. He is President of the Linden Consulting Group and holds adjunct positions in both the School of Medicine and the School of Nursing at Oregon Health and Science University.

Correspondence: Dr *Ariel Linden*, Linden Consulting Group, 6208 NE Chestnut Street, Hillsboro, OR 97124, USA.

E-mail: alinden@lindenconsulting.org