# Decision thresholds in developmental-behavioral screening: Gaining insight through system dynamics modeling

R. Christopher Sheldrick
Tufts University School of Medicine

Kee Chan
Boston University

Deborah E. Polk
University of Pittsburg

## Author Note

R. Christopher Sheldrick, Department of Pediatrics, Tufts Medical Center; Kee Chan, Department of Health Sciences, Boston University; Deborah E. Polk, Dental Public Health and Information Management, University of Pittsburgh
The project described was supported by the National Center for Research Resources, Grant Number UL1 RR025752, now at the National Center for Advancing Translational Sciences; and the National Cancer Institute, Grant Number KM1 CA156726. The authors would like to acknowledge training offered through the *Institute for Systems Science and Health* (Ann Arbor, MI, 2009), sponsored by the Office of Behavioral and Social Science Research at the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Correspondence regarding this article should be addressed to Chris Sheldrick, Tufts Medical Center, 800 Washington Street #854, Boston, MA 02111. rsheldrick@tuftsmedicalcenter.org.

## Abstract

Despite evidence for the importance of early intervention, less than one third of children with such disabilities are diagnosed before they enter school. Use of evidence-based developmental screening tools is widely recommended, but effective implementation in pediatric settings is difficult. We created an SD model to identify new strategies to detect developmental and behavioral disabilities in pediatric settings, focusing on factors that influence physicians' decision thresholds when identifying disabilities. The model was informed by the literature on medical errors and decision making, regret theory, and a prior SD model of thresholds in social policy. After testing the model for internal validity and calibrating it to known data, we conducted a series of virtual experiments to simulate interventions to improve detection of developmental and behavioral disorders among children, including introduction of a high-quality screening instrument. Results demonstrate that if one assumes that physicians adjust their decision thresholds based on feedback regarding patient outcomes, then detection rates may be improved not only by introducing screening instruments, but also by improving feedback to physicians regarding medical errors or by decreasing regret associated with false positive results (e.g., by providing convenient, non-stigmatizing referrals). The model also suggests that physicians' decision thresholds may oscillate over time, as is common in complex systems characterized by feedback with differential delays. Overall, our model demonstrates that strategies beyond use of formal screening instruments may be helpful in improving detection rates—a useful insight early in our program of research.

# Decision thresholds in developmental-behavioral screening: Gaining insight through system dynamics modeling

Causal theories are central to research, guiding everything from formulation of hypotheses to selection of variables for regression analyses. System dynamics (SD) models depict such theories graphically, but go further by specifying parameters and equations to describe relationships among variables, thus allowing for simulated "virtual experiments." If well-specified, calibrated, and validated, SD models can be powerful tools suitable for comparing competing health policies. However, small SD models can also have an important role. Because they are comparatively accessible and easy to understand, small SD models can be especially powerful tools for gaining insight into a problem (Ghaffarzadegan, Lyneis & Richardson, 2011), for example early in a research program when few data are available. Through better understanding of feedback processes, even preliminary SD models can offer insight into the implications of causal hypotheses before they are tested empirically.

We created an SD model to identify new strategies to detect developmental and behavioral disabilities, which affect up to 20% of children in the U.S. Despite evidence for the importance of early intervention, less than one third of children with such disabilities are diagnosed before they enter school (Sand et al., 2005). To improve the rate of diagnosis, therefore, organizations such as the *Council on Children with Disabilities* (2006) have recommended that pediatricians use evidence-based screening instruments (typically in the form of brief questionnaires administered to parents), and their use has increased sharply in the past decade (Radecki, Sand-Loud, O'Connor, Sharp & Olson, 2011).

However, it is not clear how pediatricians use screening instruments in practice. Many pediatricians are reluctant to follow evidence-based protocols (Lorenz et al., 2005), and many have been found to depart from expert screening recommendations for other health problems, including diabetes (Rhodes et al., 2006), obesity (Klein et al., 2010), and cardiac problems (Bensky, Covitz & DuRant, 1999). Developmental screening is typically only one element of pediatric surveillance, which also includes direct examination of the child and discussion with parents (Dworkin, 1989). Pediatricians are encouraged to use the full range of information available to them when making treatment and referral decisions. Thus, pediatricians' decisions are rarely guided solely by the results of a single screening instrument. In effect, pediatricians may be setting their own decision thresholds, rather than following those recommended for each screening test. A systematic review found that when pediatric providers identify behavioral disorders in general practice, specificity (i.e., proportion of children without disabilities who are accurately identified) typically far exceeds sensitivity (i.e., proportion of children with disabilities who are accurately identified), indicating a reluctance to commit false positive errors (Sheldrick, Merchant & Perrin, 2011). In contrast, thresholds for screening instruments typically balance sensitivity and specificity, suggesting a possible mismatch between recommendations and the realities of primary care.

## Methods

We developed an SD model of factors that influence physicians' decision thresholds. Given the preliminary nature of our data, we emphasize that our purpose was to develop theory, not make specific predictions. To facilitate closer examination of

our model, we present a detailed description of all key stocks, flows and variables in our supplementary materials.

We used three primary sources of information to define our model: (1) *a review of empirical evidence* demonstrating that physicians vary widely in their diagnosis and referral rates, even after adjusting for case mix (e.g., Franks et al., 2000), and that differences among physicians' risk perceptions are associated with an array of medical decisions (e.g., Nightingale & Grant, 1988); (2) *regret theory*, which posits that clinicians set decision thresholds based on their perceptions of the relative harm associated with false positive and false negative outcomes (Tsalatsanis, Hozo, Vickers, & Djulbegovic, 2010); and (3) a prior *SD model that describes oscillations in decision thresholds* associated with social phenomena such as the level of evidence required for police searches (Weaver & Richardson, 2006).

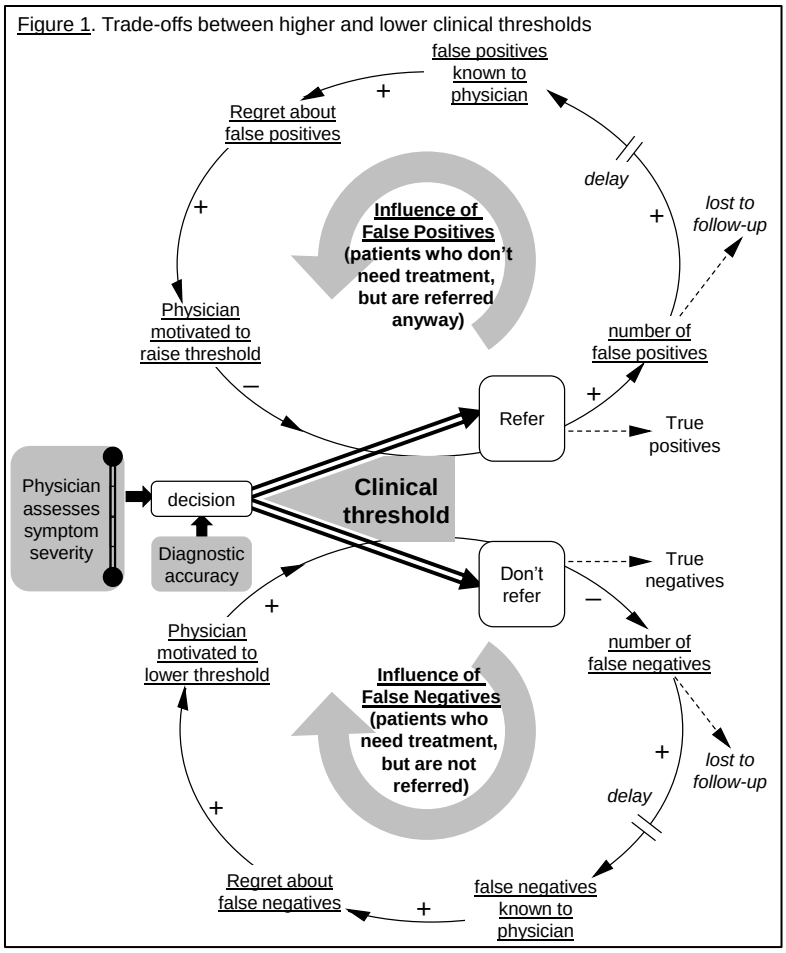
Based on these sources, we developed hypotheses that informed *model structure*. We then conducted tests of *internal validity* to ensure the model performed as expected. Next, the model was calibrated by adjusting parameters within plausible ranges until expected values of sensitivity and specificity were obtained. Finally, specific parameters were systematically altered in a series of *virtual experiments* designed to simulate interventions to improve detection rates, including introduction of a high-quality screening instrument.

**Model Structure.** In our model, physicians' initial assessments of patients are assumed to be continuous, similar to ratings on the well-known scale for physicians known as the Children's Global Assessment of Functioning (CGAS; Schaffer et al., 1983). Physicians then make referral decisions by applying a clinical threshold to this continuous

assessment. If symptom severity is found to be higher than the clinical threshold, the patient is referred (a *positive* result); otherwise, the patient is not referred (a *negative* result). In the absence of perfect diagnostic accuracy, some patients will be referred who should not be (i.e., a *false positive* result), and other patients will not be referred who should be (i.e., a *false negative* result). Our central hypothesis is that physicians adjust their decision thresholds based on their perceptions of the likelihood and impact of false positive and false negative errors, using feedback from past clinical encounters as a guide. Focusing on errors is consistent with regret theory and with previous SD models of decision thresholds (Weaver & Richardson, 2006).

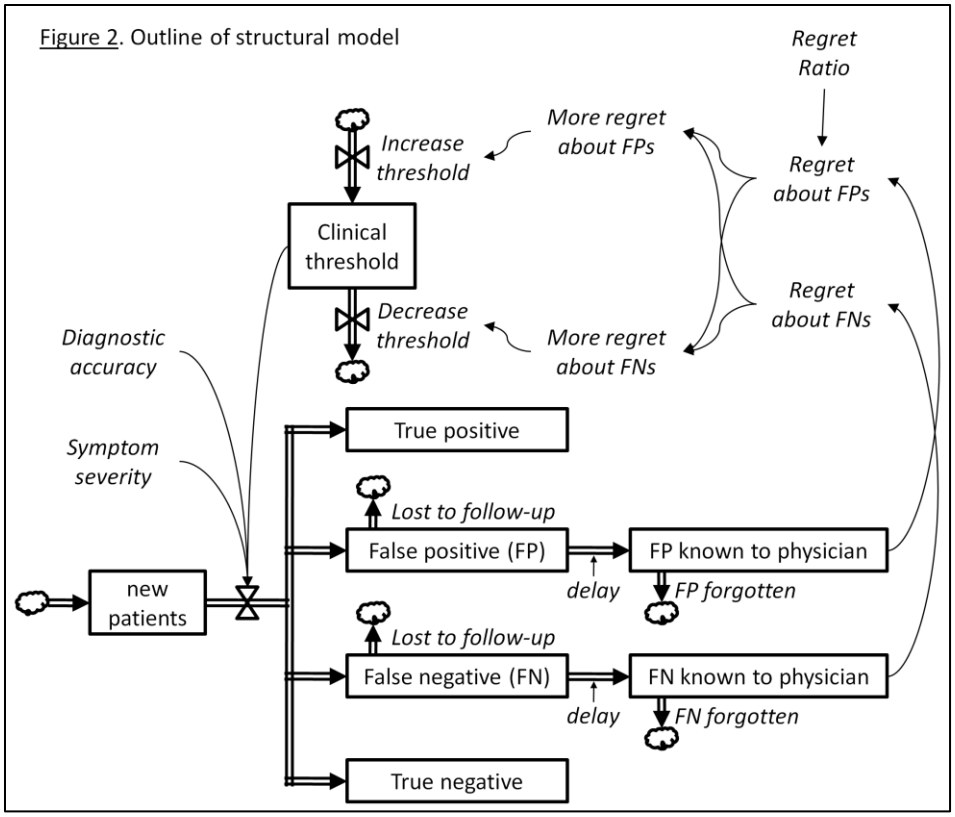
Figure 1 depicts our conceptual model. Physicians assess symptom severity with a given level of diagnostic accuracy for each of a series of patients. By applying a clinical threshold, they then make a decision to refer or not to refer for each patient. Each decision contributes to a separate feedback loop that ultimately influences the physician's clinical threshold. Examining each loop in more detail, decisions to refer or not to refer are sometimes correct

Figure 1. Trade-offs between higher and lower clinical thresholds



(leading to *true positives* and *true negatives*, respectively), but are sometimes incorrect (leading to *false positives* and *false negatives*). Errors are sometimes *lost to follow-up*, but other times become *known to the physician* after a delay. Known errors lead to *regret*, thus motivating physicians to either raise or lower their clinical thresholds to reduce the type of error for which they have more regret at that moment.

Figure 2 outlines our structural model, which rests on a standard decision tree. Assessment results (i.e., true positive, false positive, false negative or true negative) are modeled by drawing random numbers from a bivariate normal distribution in which one axis represents *symptom severity*, a second axis represents the physician's perception of symptom severity, and the correlation between the two axes (rho) represents *diagnostic accuracy*. In our model, prevalence of developmental-behavioral disorders is assumed to be 15% (Boyle et al., 2011). Therefore, if *symptom severity* falls in the top 15%, the patient is classified as having disease. If the physician's perception of symptom severity exceeds the *clinical threshold*, the patient is classified as positive;



otherwise, the patient is classified as negative.

The literature on medical errors suggests that physicians seldom receive feedback regarding patient outcomes (Redelmeler, 2005). Therefore, our model allows the proportion of each type of error that becomes known to the physician to vary between 0% and 100%; any remainder is lost to follow-up. Also based on the literature, our model assumes that knowledge of errors is often delayed (Rudolph & Morrison, 2008; Schiff, 2008).

Finally, we assume that when physicians learn of errors, they adjust their decision thresholds, thus affecting future decisions. Both false negative and false positive errors cause regret. In our model, the relative regret associated with each type of error is expressed by the *regret ratio*. For example, if regret associated with false negatives exceeds regret associated with false positives, theory suggests that physicians will lower their decision thresholds, thus identifying more positive cases and yielding fewer false negatives.

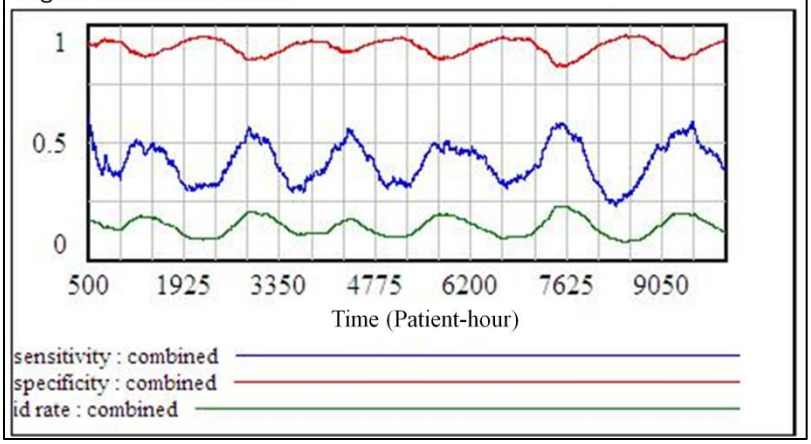
**Internal Validation.** We conducted a series of tests to ensure that our model performs as expected. We tested the model under an assumption of perfect diagnostic accuracy to ensure that the expectation of perfect sensitivity and specificity are met. We calculated the regret ratio that should yield equivalent sensitivity and specificity based on regret theory (Tsalatsanis et al., 2010), and tested whether the model behaved accordingly. Finally, we conducted tests to ensure that changes in decision thresholds were explainable given prior theory.

**Model Calibration.** For two reasons, we conducted only general calibrations to ensure that the baseline model yielded plausible results: (1) relevant time-series data on

physicians' behavior were unavailable, and (2) our stated goal was to enhance insight, not make specific predictions. Thus, we chose parameter values that are consistent with available literature and that yielded results for physician sensitivity and specificity consistent with a recent systematic review (Sheldrick et al., 2011). A table of all parameters is provided with our supplemental materials. For example, we assumed that pediatricians regret missing a true case (a false negative) three times as much as referring a patient who does not benefit (a false positive). Although we recognize that selecting a "regret ratio" of 3 is an arbitrary choice, we based it on the observation that a large percentage of pediatric referrals for mental health treatment do not result in services. Pediatricians' continued willingness to refer patients under such conditions suggests acceptance of a number of false positives for every child who ultimately receives services. Furthermore, we assumed that physicians' knowledge of false positive results, although imperfect, is greater than their knowledge of false negative results. Because there are seldom formal systems to detect missed mental health diagnoses and to report such errors back to the pediatrician, knowledge of false negatives is likely to be rare. In contrast, pediatricians are more likely to learn that a referred patient was found to be ineligible for services. We adjusted loss-to-follow-up parameters accordingly. Our final baseline model yielded an average sensitivity of 42% and an average specificity of 90%, which are generally consistent with previous research (Sheldrick et al., 2011). However, feedback delays inherent in our model's structure yielded oscillation in the clinical threshold and in these values (see Figure 3).

**Evidence of external validity.** After developing and calibrating our model, we searched the pediatric literature for recently published studies that report data relevant to our model.

Figure 3. Baseline model




**Virtual Experiments.** To examine the effect of parameter changes on sensitivity and specificity, we conducted several virtual experiments. First, we simulated the effect of introducing a high-quality screening instrument by altering the diagnostic accuracy parameter (rho) from .65 (which describes a Receiver Operating Characteristics [ROC] curve that includes sensitivity and specificity equal to 75%) to .85 (which describes an ROC curve that includes sensitivity and specificity equal to 85%). Because most policy statements on pediatric screening are concerned with increasing detection, we then altered the following parameters, one at a time, to produce equivalent changes in sensitivity:

- Regret ratio. Various factors may alter physicians' regret regarding false negative versus a false positive errors (i.e., the regret ratio). For example, referrals for mental health services can be time-consuming for both providers and patients. In addition, concern about stigma sometimes makes providers reluctant to make mental health referrals. Convenient and non-stigmatizing follow-up services, possibly through collaborative, co-located mental health care, may reduce patients' burden, making false-positive results more tolerable. Reducing

physicians' regret for false positives may make lower clinical thresholds more tolerable.

- Loss-to-follow up for false negatives. Some investigators recommend formal systems to provide physicians with systematic feedback to improve patients' mental health outcomes (Bickman, 2008). We simulated such a solution by reducing loss-to-follow-up for false negatives. Because regret is dependent on knowledge of errors, greater knowledge of false negative errors may make high clinical thresholds less tolerable.

Finally, we tested the combined effect of all three parameter changes.

## Results

As stated above, the baseline model simulated pediatricians' identification of children with developmental and behavioral disorders without the use of screening instruments. This model yielded sensitivity=42% and specificity=90%, values that are plausible in light of a recent systematic review (Sheldrick et al., 2011). Consistent with previous research (Weaver & Richardson, 2006), our model displayed oscillations over time in decision thresholds, sensitivity, and specificity.

Parameter values for the baseline model and virtual experiments are presented in Table 1. Figure 4 presents results of virtual experiments. Note that the

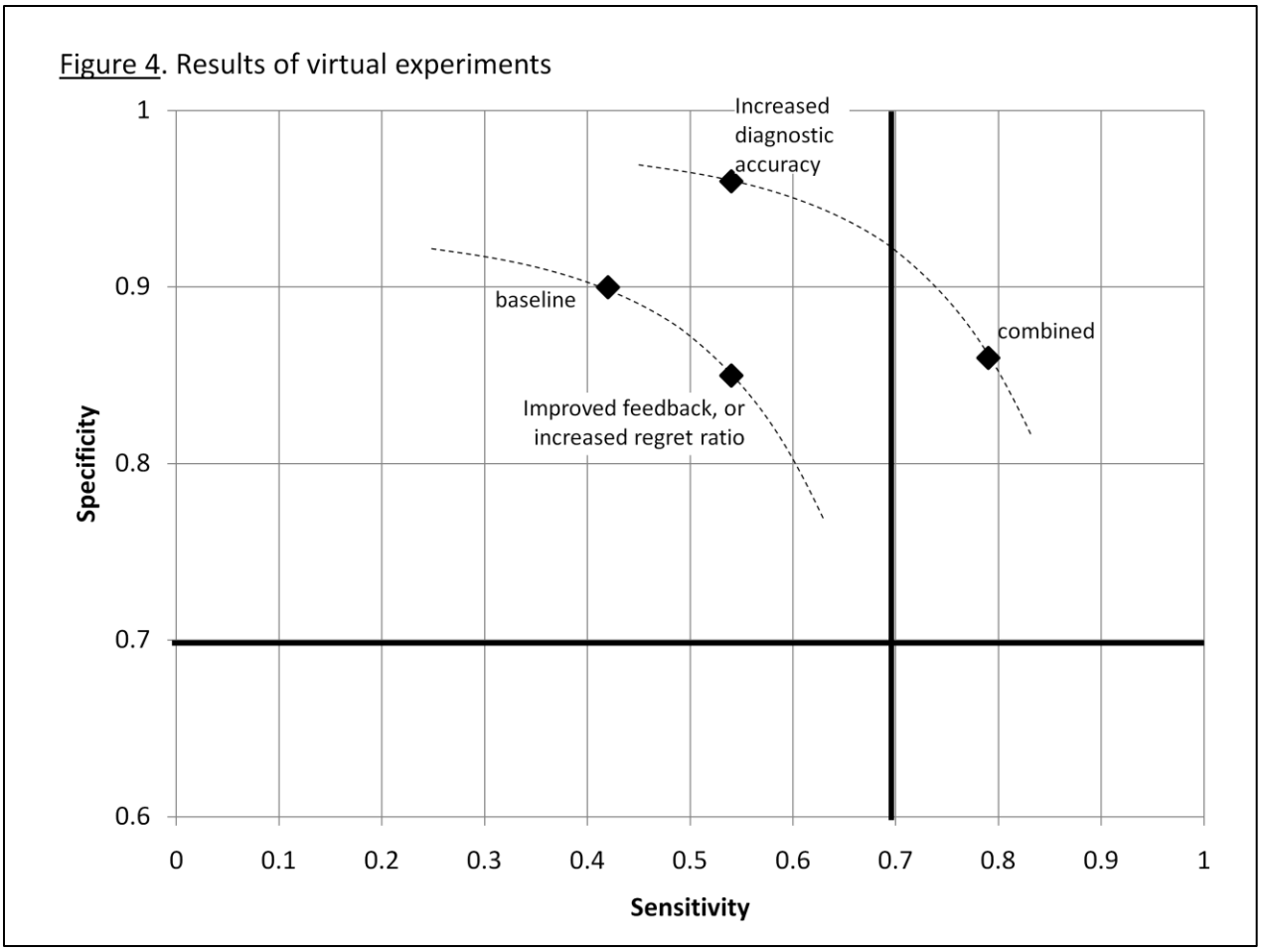
Table 1. Parameter values for virtual experiments

| model name | Model parameters | | | |
|---|---|---|---|---|
| | False positives (FP) lost to follow-up | False negatives (FN) lost to follow-up | Diagnostic accuracy (rho) | Regret ratio (FN/FP) |
| Baseline | 33% | 80% | 0.65 | 3 |
| Increased diagnostic accuracy | 33% | 80% | **0.85** | 3 |
| Improved feedback | 33% | **60%** | 0.65 | 3 |
| Increased regret ratio | 33% | 80% | 0.65 | **6** |
| Combined | 33% | **60%** | **0.85** | **6** |

dark lines are placed at 70% sensitivity and 70% specificity, which are widely considered to be minimum standards for developmental-behavioral screening instruments. Thus, points falling in the upper-right hand quadrant of Figure 4 can be considered to display acceptable sensitivity and specificity.

In our first virtual experiment, we increased diagnostic accuracy to simulate pediatricians' use of validated screening instruments. Although 85% sensitivity and specificity were possible under this condition, the model reached equilibrium at sensitivity = 54% and specificity = 96%, reflecting a different point on the same ROC curve (see "increased diagnostic accuracy" in Figure 4). Similar improvements in sensitivity were achieved by our two other simulated interventions: (1) reducing loss-to-follow-up for false negatives from 80% to 60%, and (2) increasing the regret ratio from 3

Figure 4. Results of virtual experiments



to 6 (see “improved feedback or increased regret ratio” in Figure 4). In each case, the decision threshold was lowered, but overall diagnostic accuracy was unaffected. For these latter two interventions, gains in sensitivity were accompanied by loss of specificity to 85%.

Changing all three parameters simulated a combined intervention in which diagnostic accuracy was increased, regret about false-positives was decreased, and feedback about false negatives was improved. In this simulation, sensitivity and specificity were estimated at 79% and 86%, respectively (see “combined” in Figure 4). Together, these simulations suggest that strategies that go beyond encouraging pediatricians to use formal screening tools may be helpful in improving detection rates.

We conducted a final literature search to identify recent studies that could provide tests of the external validity of our model. Our search revealed a recent pragmatic clinical trial that randomly assigned children to either receive standard pediatric care or standard care plus developmental screening. Among children identified with delays, 58% were referred for services (Guevara et al., 2013). In virtual experiment #1 of our model, a screener is implemented that is capable of 85% sensitivity, but the model reaches equilibrium at 54% sensitivity. Thus, our model predicts that 63.5% (i.e., 54/85) of children with developmental disabilities will be referred if a developmental screening program is implemented.

## Discussion

Results demonstrate that if one assumes that physicians adjust their decision thresholds based on feedback regarding patient outcomes, then multiple strategies to improve sensitivity are possible. Improving diagnostic accuracy (e.g., by introducing

formal screening) will improve sensitivity and specificity, but the differential between the two will remain. In contrast, improving feedback by reducing loss to follow-up or decreasing regret associated with false positive results (e.g., by providing convenient, non-stigmatizing referrals through co-located collaborative care) increases sensitivity by changing decision thresholds, thus decreasing the differential between sensitivity and specificity. Combining all three interventions yielded the most positive results, with sensitivity (79%) and specificity (86%) both within the range that is typically deemed acceptable in developmental-behavioral screening.

An additional insight is that decision thresholds may oscillate over time, even given stable parameters. Such behavior is common in complex systems characterized by feedback with differential delays (Sterman, 2000), and oscillations have long been observed in decision thresholds in public policy (Weaver & Richardson, 2011). However, we know of no investigations of oscillations in physicians' decision thresholds.

We highlight two limitations to our model. First, data to support precise parameter estimates were unavailable. Thus, we conducted a series of additional analyses with varying parameter estimates and found no substantive difference in results. We also emphasize that our model is designed to enhance insight—not make specific predictions. Although our model predicts results that are generally consistent with a recent pragmatic clinical trial of developmental screening (Guevara et al., 2013), caution should be exercised in interpreting specific predictions made by this model. Second, every model—ours included—is limited by its scope. Simplifying assumptions are necessary to produce a workable, understandable model. In particular, several exogenous variables could themselves be the focus of additional models. For example,

physicians' regret is likely to be informed by a range of factors, including personal (e.g., personality, training), interpersonal (e.g., influence from other physicians), and environmental (e.g., quality of mental health resources, chance of audit or lawsuit), each of which could be the focus of a more detailed model. Likewise, our model does not consider downstream effects of physicians' decisions on other service providers.

Nevertheless, we believe our model offers unique insights to guide future research. In particular, we recommend further investigation of physicians' decision-making, including collection of longitudinal, patient-level data. We also recommend study of novel interventions to improve physicians' identification of developmental-behavioral disorders, including systematic feedback systems, as well as strategies to reduce pediatricians' perceived barriers to patient referral.

## References

Bensky, A.S. Covitz, W. & DuRant, R.H. (1999). Primary care physicians' use of screening echocardiography. *Pediatrics. 103,* e40.

Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the Academy of Child and Adolescent Psychiatry, 47*, 1114-1119.

Boyle, C.A., Boulet, S., Scheive, L.A., Cohen, R.A., Blumberg, S.J., Yeargin-Allsop, M., Visser, S., Kogan, M.D. (2011). Trends in the prevalence of developmental disabilities in US children, 1997-2008. *Pediatrics, 127*, 1034-1042.

Council on Children with Disabilities [Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children with Special Needs Project Advisory Committee]. (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics, 118*, 405-420.

Dworkin PH (1989). British and American recommendations for developmental monitoring: the role of surveillance. *Pediatrics. 84*:1000–1010.

Franks, P., Williams, G.C., Zwanziger, J., Mooney, C. & Sorbero, M. (2000). Why do physicians vary so widely in their referral rates? *Journal of General Internal Medicine, 15*, 163-168.

Ghaffarzadegan N., Lyneis, J., Richardson, G.P. (2011). How small system dynamics models can help the public policy process. *System Dynamics Review, 27*, 22-44.

Guevara, J.P., Gerdes, M., Localio, R., Huang, Y.V., Pinto-Martin, J., Minkovitz, C.S., Hsu, D., Kyriakou, L., Baglivo, S., Kavanaugh, J., Pati, S. (2013). Effectiveness of developmental screening in an urban setting. *Pediatrics, 131*, 30-27.

Klein, J.D., Sesselberg, T.S., Johnson, M.S., O'Connor, K.G., Cook, S., Coon, M., Homer, C., Krebs, N. & Washington, R. (2010). Adoption of body mass index guidelines for screening and counseling in pediatric practice. *Pediatrics, 125,* 265-272.

Lorenz, K.A., Ryan, G.W., Morton, S.C., Chan, K.S., Wang, S. & Shekelle, P.G. (2005). A qualitative examination of primary care providers' and physician managers' uses and views of research evidence. *International Journal for Quality in Health Care, 17,* 409-414.

Nightingale, S.D. & Grant, M. (1988). Risk preference and decision making in critical care situations. *Chest, 93*, 684-687.

Radecki, L., Sand-Loud, N., O'Connnor, K.G., Sharp, S. & Olson, L.M. (2011). Trends in the use of standardized tools for developmental screening in early childhood: 2002-2009. *Pediatrics, 128*, 14-19

Redelmeler, D. (2005). The cognitive psychology of missed diagnoses. *Annals of Internal Medicine, 142*, 115-120,

Rhodes, E.T., Finkelstein, J.A., Marshall, R., Allen, C., Gillman, M.W. & Ludwig, D.S. (2006). Screening for type 2 diabetes mellitus in children and adolescents: attitudes, barriers, and practices among pediatric clinicians. *Ambulatory Pediatrics. 6,* 110-114.

Rothman, K.J., Greenland, S., Poole, C. & Lash, T.L. (2008). Causation and causal inference. In K.J. Rothman, S. Greenland & T.L. Lash (Eds.) *Modern Epidemiology, 3rd Edition*. Lippicott Williams & Wilkins: Philadelphia, PA.

Sand, N., Silverstein, M., Glascoe, F.P., Gupta, V.B., Tonniges, T.P. & O'Connor, K.G. (2005). Pediatricians' reported practices regarding developmental screening: Do guidelines work? Do they help? *Pediatrics*, *116*, 174-179.

Schaffer D., Gould, M.S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., Aluwahlia, S. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry, 40*, 1228-1231.

Schiff, G. (2008) Minimizing diagnostic error: The importance of follow-up and feedback. *The American Journal of Medicine, 121*, S38-42

Sheldrick, R.C., Merchant, S., Perrin, E.C. (2011). Identification of developmental-behavioral disorders in primary care: A systematic review. *Pediatrics, 128,* 356-363.

Sheldrick, R.C. & Perrin, E.C. (in press). Evidence-based milestones for surveillance of cognitive, language and motor development. *Academic Pediatrics*. DOI: 10.1016/j.acap.2013.07.001

Sterman, J.D. (2000) Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill/Irwin: New York.

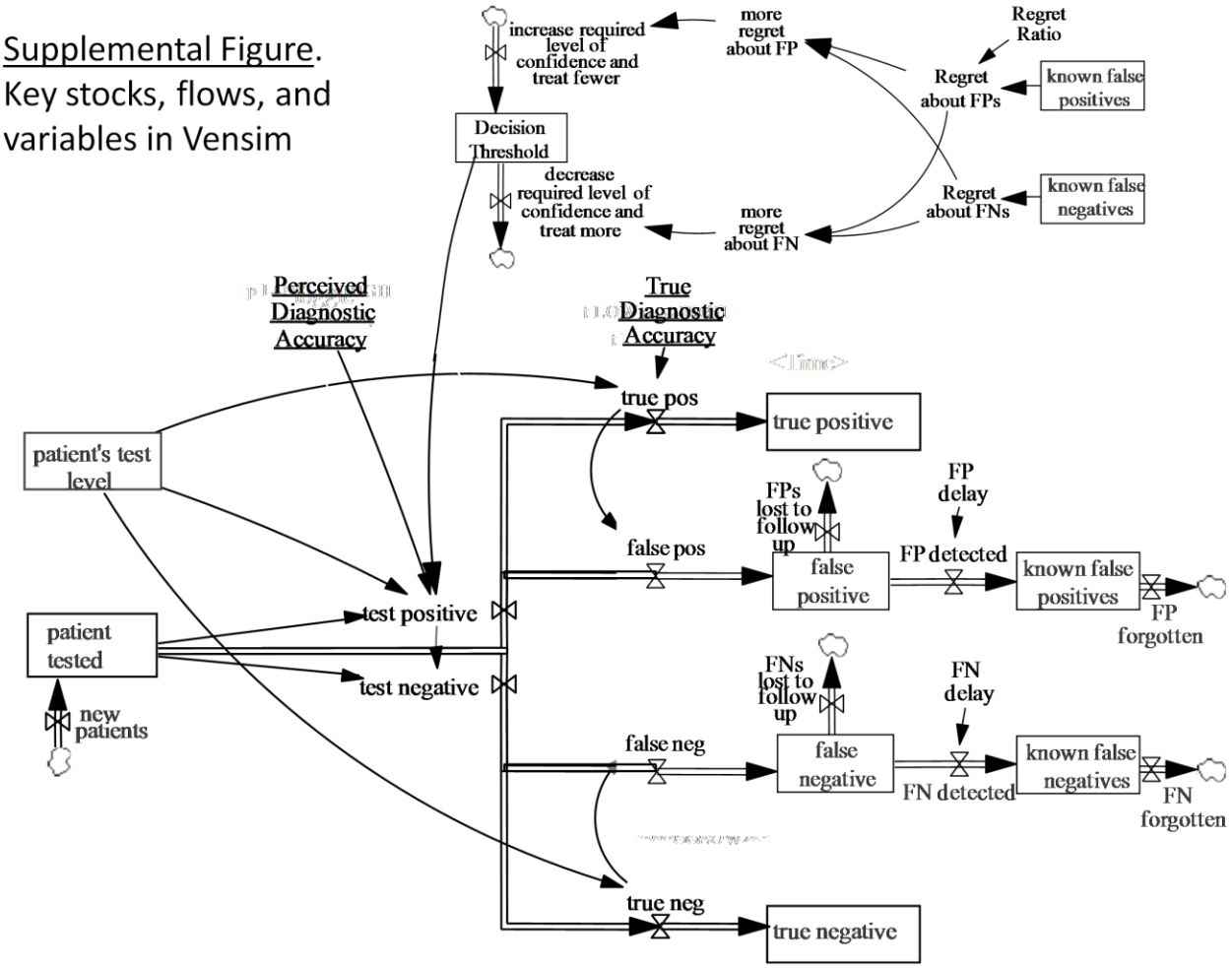
Tsalatsanis, A. **,** Hozo, I.**,** Vickers, A. & Djulbegovic, B. (2010). A regret theory approach to decision curve analysis: A novel method for eliciting decision makers' preferences and decision-making. *BMC Medical Informatics and Decision Making, 10*, e51.

Weaver, E.A. & Richardson, G. (2006). Threshold setting and the cycling of a decision threshold. *System Dynamics Review, 22*, 1-26.

**Supplemental materials**.

Below we present a figure depicting key stocks, flows, and variables as specified in Vensim (see Supplemental Figure), along with descriptions of the same (see Supplemental Table).

Supplemental Figure. Key stocks, flows, and variables in Vensim



Supplemental Table. Values and equations for model variables

| Exogenous Variables | Value | Description |
|---|---|---|
| Initial time | 0 patient-hours | |
| Time step | 1 patient-hour | |
| Final time | 10,000 patient-hours | |
| patient's test level | variable | An observation is drawn from a bivariate normal distribution is generated per each patient-hour. Patient's test level is one coordinate of this observation. |
| Perceived Diagnostic Accuracy | Four look-up tables; initial value = low accuracy | each table is based on a bivariate normal distribution modeling the association between "patient's test level" and true test level. Low, moderate, high and perfect accuracy are modeled by varying the correlation between these two values (rho = .65, .75, .85 and 1.0 for low, moderate, high and perfect accuracy, respectively). |
| True Diagnostic Accuracy | " | Coded as above; for purposes of this model, perceived and true accuracy were considered to be equal |
| Has Disease | | Uses input from *True Diagnostic Accuracy.* Generates a random number from a uniform distribution and compares it to the probability of disease. If greater, then *Has Disease* = 1; if less; *Has Disease* = 0. Note that for cases identified as positive, the probability of disease increases with diagnostic accuracy. |
| percent FP | initial value = 0.00 | initially assumes 0% of false positives are lost to follow-up |
| percent FN | initial value = 0.00 | initially assumes 0% of false positives are lost to follow-up |
| FP delay | initial value = 1 patient-hour | This parameter is the denominator of "FP detected." Initial value therefore assumes no effect. Greater values indicate more delay. Measured in patient-hours, with an assumption of 20 patient-hours per day. |
| FN delay | initial value = 1 | This parameter is the denominator of "FN detected." Initial value therefore |

| | | |
|---|---|---|
| | patient-hour | assumes no effect. Greater values indicate more delay. Measured in patient-hours, with an assumption of 20 patient-hours per day. |
| Regret Ratio | initial value = 1 | equals (regret FN)/(regret FP). Is generally equivalent to (1-Pt)/Pt, where Pt is a decision or probability threshold; assumes that regret about missing a diagnosis (FN) is 1.0 times greater than regret about an incorrect diagnosis (FP); as concern over FP rises, this ratio falls. |
| **Endogenous Variables** | **Type** | |
| new patients | flow | assumes 1 new patient per patient-hour |
| patient tested | stock | *=new patients - test positive - test negative* |
| test positive | flow | uses input from *Perceived Diagnostic Accuracy.* If test level exceed threshold, then *test positive = patient tested* |
| test negative | flow | dependent on *test positive;* if patient does not test positive, then *test negative = patient tested* |
| true pos | flow | uses input from *Has Disease.* If patient has disease, then true pos = test positive. |
| false pos | flow | dependent on *true pos.* If *true pos = 0* , then *false pos = test positive.* |
| false neg | flow | uses input from *Has Disease.* If patient has disease, then false neg = test positive. |
| true neg | flow | dependent on *false neg.* If *false neg = 0* , then *true neg = test positive.* |
| true positive | stock | *=true pos-LTF TP-TP detected* |
| false positive | stock | *=false pos-FP detected-FPs lost to followup* |
| false negative | stock | *=false neg-FN detected-FNs lost to followup* |
| true negative | stock | *=true neg-LTF TN-TN detected* |
| TP detected | flow | *=true positive* |
| FP detected | flow | *=false positive / FP delay* |
| FN detected | flow | *=false negative / FN delay* |
| TN detected | flow | *=true negative* |
| known true positives | stock | *=TP detected-TP forgotten* |
| known false positives | stock | *=FP detected-FP forgotten* |
| known false negatives | stock | *=FN detected-FN forgotten* |
| known true negatives | stock | *=TN detected-TN forgotten* |
| TP forgotten | flow | assumes regret from a known *true positive* dissipates over two years (divides by 2*365*20) |

| | | |
|---|---|---|
| FP forgotten | flow | assumes regret from a known *false positives* dissipates over two years (divides by 2*365*20) |
| FN forgotten | flow | assumes regret from a known *false negatives* dissipates over two years (divides by 2*365*20) |
| TN forgotten | flow | assumes regret from a known *true negatives* dissipates over two years (divides by 2*365*20) |
| Regret about FNs | variable | *=known false negatives* |
| Regret about FPs | variable | *=known false positives/Regret Ratio* |
| more regret about FN | variable | if Regret about FNs > Regret about FPs, then .0005; else 0 |
| more regret about FP | variable | if Regret about FPs > Regret about FNs, then .0005; else 0 |
| increase required level of confidence and treat fewer | flow | *=more regret about FP* |
| decrease required level of confidence and treat more | flow | *=more regret about FN* |
| Decision Threshold | stock | *=increase required level of confidence and treat fewer-decrease required level of confidence and treat more* |