Early Identification of Short-Term Disability Claimants Who Exhaust Their Benefits and Transfer to Long-Term Disability Insurance

Kara Contreary
Mathematica Policy Research

Yonatan Ben-Shalom
Mathematica Policy Research

Brian Gifford
Integrated Benefits Institute

5th Annual Meeting of the Disability Research Consortium

August 2, 2017
Washington, DC

The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Disability Research Consortium. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA or any agency of the federal government. Neither the U.S. government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation, or favoring by the U.S. government or any agency thereof.

1. Introduction

Short-term disability insurance (STDI) pays partial wage replacements to employees who are temporarily unable to work due to illnesses or injuries that are not work related. STDI coverage has grown in recent years; close to 40 percent of private sector workers were covered in 2014 (Monaco 2015). Most policies replace wages for a fixed period of time, with a median coverage length of six months (although some plans cover a year or more). Because lost wages are replaced only partially—the median replacement rate is 60 percent (Monaco 2015)—employees have an incentive to return to work. STDI claimants who are unable to return to work before their benefits expire may be at higher risk of job loss, receipt of long-term disability insurance (LTDI) benefits, or receipt of Social Security Disability Insurance (SSDI) benefits.

Although previous research sheds light on how LTDI plan characteristics influence take-up and duration of claims (Autor et al. 2014), relatively little is known about the factors that influence STDI duration or about the transition from STDI to long-term or Social Security disability benefits. An STDI claim can be an early point of identification of workers who have medical conditions but could, with adequate support, remain in the work force. However, careful timing and targeting of early intervention is critical to efficiency; some workers may return to work without intervention, while others may not benefit from intervention (Stapleton et al. 2015). In both cases, spending resources on early intervention would not result in improved labor force retention.

In this paper, we explore two primary research questions about exhausting STDI:

Which observable factors help predict exhaustion of STDI benefits and transfers to LTDI?

Can waiting for some claims to resolve without intervention improve efficiency of targeting individuals for early intervention aimed at helping them remain in the work force?

2. Methods

Data

We use Integrated Benefits Institute (IBI) Health and Productivity Benchmarking Data from 2011 through 2015. The data used in the current analysis include 1.1 million closed STDI claims from 15,212 small, medium, and large businesses associated with 9 disability insurance carriers and third-party leave administrators. The claims data are collected for leave administration purposes by the carriers and administrators, and include information on claim outcomes (reached maximum duration, transition to LTDI); claimant characteristics (age, sex, wages, primary diagnosis, census division); employer characteristics (industry, size), and insurance plan design characteristics (elimination period, maximum benefit duration). Our outcomes of interest are exhaustion of the STDI benefit and transition to LTDI. Information on LTDI transition is available only for a subset of claims. In this summary we focus on exhaustion of the STDI benefit, but the results were qualitatively similar for LTDI transition.

Analysis

After calculating summary statistics for the primary outcomes, we estimated logit regression models with the outcomes of interest as the dependent variables, and generated predicted probabilities of exhausting STDI for each claim. To avoid overfitting, we used a split sample approach, randomly dividing our sample into a modeling half and a validating half. We used the first half in the regressions, then used the estimated coefficients to generate predicted probabilities for the second half. We then constructed receiver operating characteristic (ROC) curves, which illustrate for each model the tradeoff between sensitivity and specificity. Specificity is the true positive rate, or likelihood of correctly identifying claims that will exhaust STDI benefits. Sensitivity is the true negative rate, or likelihood of correctly identifying claims that will not exhaust STDI benefits. As the probability threshold increases, fewer claims are flagged as high risk, increasing the specificity but decreasing the sensitivity.

Our sample included claims with varying maximum benefit durations—13 weeks, 26 weeks, and 52 weeks—which we analyzed separately.[1] We assessed the predictive accuracy of our models using the area under the ROC curve (AUC). In general, a higher AUC indicates better predictive accuracy—for example, higher sensitivity for a given level of specificity, or vice versa. However, the AUC gives no guidance regarding the optimal probability threshold; that is, the predicted probability above which to flag claimants as being at "high risk" of exhausting benefits. For that we used Youden's index, the point on the ROC curve that maximizes the sum of sensitivity and specificity (Youden 1950).

We examined three models incorporating varying pieces of information about the claims: (1) age and sex only; (2) the major diagnosis category only (of the claim's primary diagnosis; for example, digestive system disorders); and (3) a richer set of information: age, sex, diagnosis category, weekly wage, census division, employer industry, employer size, and policy elimination period. Comparing results across the three models showed to what extent modeling improves the accuracy of targeting over simpler decision rules. We analyzed the three models using the full sample of claims and over time, sequentially eliminating claims that resolved within 2 weeks, 4 weeks, 6 weeks, and 12 weeks.[2] Comparing across claim durations shows the potential efficiency gains of waiting to allow some claims to resolve on their own.

3. Results

We present results for claims with a maximum benefit duration of 26 weeks, which made up 73 percent of our sample. Results were similar in the 13- and 52-week subgroups.

Factors associated with exhaustion of benefits

The individual characteristics most strongly associated with exhaustion of STDI benefits are age and diagnosis category. The probability of exhausting STDI benefits increases nearly linearly with age, reaching 6 percentage points higher among individuals ages 55 and over than among individuals ages 18 to 24. By far the diagnosis most strongly associated with STDI benefit exhaustion is cancer. Claims with a diagnosis in the category of malignant neoplasms are 11 percentage points more likely to reach maximum duration than claims for "other illnesses"

[1] A person whose condition requires seven months of leave would show up in our data as exhausting STDI benefits if that person's plan had a maximum duration of 13 weeks, but as not exhausting STDI benefits with a plan of 52 weeks. Indeed, the likelihood of exhausting STDI benefits decreases with maximum benefit duration, from 16 percent among claims with a maximum duration of 13 weeks to 4 percent among those with a maximum duration of 52 weeks. By keeping these subgroups separate, we are able to eliminate variation in likelihood of exhausting STDI that is explained not by individual or employer characteristics but by the fact that benefits of shorter duration are easier to exhaust.

[2] We did not evaluate a 12-week minimum claim duration for the 13-week maximum benefit duration subsample.

(our reference category). Back pain and mental health disorders, which together constitute a large share of SSDI awards, are also positively associated with exhausting benefits: diagnoses of "intervertebral disc disorder" and "other back disease" have a 6 and 3 percentage point higher probability of exhausting benefits, respectively; depression and PTSD are associated with a 4 percentage point higher probability of STDI exhaustion. Weekly wage is not associated with probability of exhausting STDI benefits. Sex is statistically significantly associated with benefit exhaustion but the magnitude of the association is small: women are 0.7 percentage points less likely than men to exhaust STDI benefits.

Employer size is not associated with STDI exhaustion, but industry strongly is: employment in agriculture, mining, construction, transportation, and utilities is associated with a 3 to 6 percentage point higher probability of exhausting STDI benefits than any other industry category.[3] Finally, the elimination period of the STDI insurance plan, which is the length of time between filing a claim and starting to receive wage replacements, is positively associated with STDI benefit exhaustion; claims that have elimination periods longer than two weeks are associated with a nearly 5 percentage point higher likelihood of exhausting benefits than claims that have no elimination period. This is likely due to censoring of disability claims that resolve quickly on their own.

Predicting STDI exhaustion

Benefits of waiting for claims to resolve on their own

Figure 1 shows ROC curves and Youden's indexes for the sample of claims with a maximum benefit duration of 26 weeks, at 0 weeks minimum claim duration (panel A) and 6 weeks minimum claim duration (panel B). In both cases, the full model has the highest AUC, followed by the diagnosis-only model, and then the age-and-sex-only model. Somewhat surprisingly, the AUC falls with claim duration. In week 0, at the probability threshold indicated by Youden's index, sensitivity is 65.4 percent, specificity is 63.1 percent, and the probability threshold for "high risk" is 0.08. In week 6, at Youden's index, sensitivity is 60.1 percent, specificity is 62.5 percent, and the probability threshold for "high risk" is 0.17. Table 1 shows the predictive accuracy metrics for the full model at minimum claim durations of 0, 2, 4, 6, and 12 weeks.

The results in Table 1 show that, although predictive accuracy does not improve with claims duration (AUC and Youden's index both decline over time), efficiency of targeting does improve. By week 6, compared to week 0, the model would flag half the number of claims as "high risk" and double the percent of true positives out of those flagged. Much of the benefit of waiting comes from the relatively rapid attrition of claims from the sample. By six weeks, almost 50 percent of claims resolved without intervention.[4] At least some of those were likely among the claims that would have been targeted by the model in week 0. Therefore, waiting serves the function of eliminating false positives from the set of targeted claims. However, the number of

[3] Other industries included manufacturing; wholesale and retail trade; finance, insurance, and real estate; services; and the public sector.

[4] Usual practice may in fact include some efforts on the parts of insurers and/or employers to resolve STDI claims. In the context of discussions around early intervention, we assume that proposed interventions would be in addition to business as usual, and that the attrition of claims from the sample over time forms the baseline resolution rate.

true positives (sixth column) declines only slightly relative to the number of claims targeted; waiting does not appear to greatly increase the false negative rate.

Figure 1. ROC curves and Youden's index, maximum duration of 26 weeks

Panel A: Minimum claim duration of 0 weeks Panel B: Minimum claim duration of 6 weeks

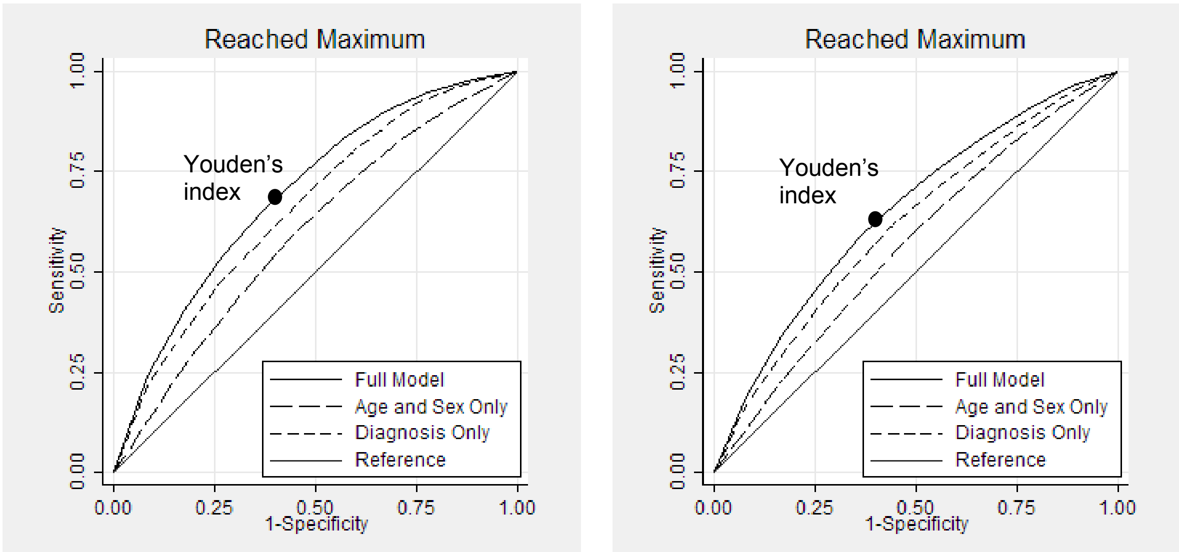

Table 1. Predictive accuracy of full model, maximum duration of 26 weeks

Minimum claim duration (weeks)	Number of claims	Area under the curve	Probability threshold (at Youden's index)	Number of claims targeted	Number of true positives	Percent of flagged that are true positives
0	451,370	0.697	0.079	181,501	20,825	11.47
2	374,770	0.682	0.098	150,153	20,080	13.37
4	290,940	0.667	0.130	118,119	19,448	16.46
6	226,270	0.651	0.171	94,429	18,930	20.05
12	129,810	0.628	0.348	60,325	18,133	30.06

Benefits of modeling versus attrition alone

To understand the advantages of modeling STDI benefit exhaustion, it is useful to compare the results from using a predictive model to the option of attrition alone, which involves waiting a number of weeks to allow claims to resolve, then targeting all claims that survive to that duration. Table 2 shows that at a claim duration of six weeks, modeling results in targeting less than half the number of claims compared to attrition alone. The cost of modeling is that it captures 60 percent rather than 100 percent of claims that would exhaust STDI benefits under usual practice (18,930 vs. 31,498 claims). However, this cost may be worthwhile because the number of false positives is reduce by over 60 percent (from 194,772 to 75,499 claims).

Table 2. Predictive accuracy of full model versus attrition alone, 6 weeks

Model	Number of claims	Number of claims targeted	Number of true positives	Number of false positive
Attrition alone	226,270	226,270	31,498	194,772
Full model	226,270	94,429	18,930	75,499

4. Discussion

Early intervention efforts require careful timing and targeting in order to efficiently support workers who are at risk for exiting the labor force. We present a basic model that demonstrates that the efficiency of early intervention efforts can be improved by waiting and using observable characteristics to model the likelihood of exhausting STDI benefits. Waiting allows claims that will resolve without intervention to do so, and modeling narrows the target population, reducing the costs of intervention. Our findings have implications for early intervention programs aimed at helping workers with medical conditions to remain in the labor force. To maximize the efficiency of intervention efforts, policymakers can best direct limited resources by strategically waiting to allow some claims to resolve without intervention, then targeting efforts at individuals who are at highest risk of exiting the labor force based on observable characteristics.

The strength of this approach is that it uses a large data set of private STDI claims and commonly used and well-understood predictive accuracy metrics. However, our work has two main limitations. First, our specification of the full model is very basic, with all terms entering linearly; a more complex model may improve predictive accuracy. In future work, we will test whether an optimized model specification substantially improves predictive accuracy. Another option is to use machine learning to identify constellations of characteristics that are highly predictive of STDI exhaustion.

Second, both AUC and Youden’s index treat sensitivity and specificity as equally important components of predictive accuracy. However, in the context of early intervention to promote labor force retention, it is unlikely that false positives (providing treatment to someone who does not need it or will not benefit from it) and false negatives (failing to intervene with someone who could benefit) are equally important. Which one is weighted higher depends on a number of factors, including the cost of the intervention, its effectiveness over time, and the costs associated with progressing to LTDI or SSDI. Nevertheless, we use these measures here because of their prevalence in the literature and to focus attention on how predictions are improved by waiting and by using all available information. In future work, we will relax the assumption of equal preference for sensitivity and specificity.

References

Autor, David, Mark Duggan, and Jonathan Gruber. “Moral Hazard and Claims Deterrence in Private Disability Insurance.” *American Economic Journal: Applied Economics,* vol. 6, no. 4, October 2014, pp. 110–141.

Monaco, Kristen. “Disability Insurance Plans: Trends in Employee Access and Employer Costs.” *Beyond the Numbers: Pay and Benefits,* vol. 4, no. 4, U.S. Bureau of Labor Statistics, February 2015. Available at https://www.bls.gov/opub/btn/volume-4/disability-insurance-plans.htm.

Stapleton, David, Robert Burns, Winthrop Cashdollar, Benjamin Doornink, Mary Harris, Robert Anfield, Brian Gifford, and Kevin Ufier. "Targeting Early Intervention to Workers Who Need Help to Stay in the Labor Force." Washington, DC: Mathematica Policy Research, October 2015.

Youden, William J. "Index for Rating Diagnostic Tests." *Cancer*, vol. 3, no. 1, 1950: pp. 32–35.