

Improving Prescription Drug Risk Assessment Tools

ASSessment tools that assign risk scores to insureds—or potential insureds—based on their prescription histories are increasingly used by actuaries and others within health insurance organizations to assess future health risk. Prescription drug and diagnosis-based risk assessment tools (the latter, a category of tools that key off diagnosis codes and are often used in conjunction with or instead of prescription drug tools) are used in areas as diverse as new business and renewal underwriting for small and medium-size groups, utilization and disease management, trend analysis, comparisons between blocks of business, and provider compensation.

Prescription drug risk assessment tools have been built by statisticians and have confirmed statistical validity. (An analysis of their validity can be found at <http://www.soa.org/files/pdf/risk-assessmentc.pdf>.) Yet when potential early adopter underwriters and medical directors have looked at combining risk assessment tools with the prescription histories available from prescription history profile services such as Milliman Intelliscript and Ingenix Medpoint, they more often than not have concluded that the risk assessment scores do not adequately reflect their applicants' potential risk, even with respect to producing a relative ranking of applicants.

As actuaries who buy and use prescription drug risk assessment tools, we should be very concerned about the unfavorable evaluation of these tools by underwriters and medical directors. If risk isn't being well-assessed at the individual level, then it's not being well-assessed at the group or block-of-business level.

Today's Tools

With respect to the underlying algorithms, today's leading commercial-population prescription drug risk assessment tools are quite similar to one another. In generically describing the algorithms, I will not be quite true to either as regards vocabulary or technical details.

The algorithms start by assigning drugs filled within the assessment period to drug categories. Although there are

exceptions with combination drugs, each drug is typically assigned to only one category.

The assessment period is usually one year but may be longer or shorter depending on the data available and the purpose for which they are being used. Intuitively, the length of the assessment period should affect the predictive value, but I'm not aware of formal studies on the statistical impact of using periods of other than one year.

The drug categories definitions are developed by the tool vendors based on medical advice and tests for statistical validity. A given drug category may correspond to either a specific chemical composition/biological pathway or therapeutic use (the condition(s) that the drugs commonly treat).

For example, broad-spectrum oral antibiotics all may be put into a single category, while an antibiotic commonly used to treat infections associated with one particular condition may be in another category, and a particularly powerful "third-generation" antibiotic may be in yet another category.

After the drugs are assigned to categories, each category is assigned a 0 or a 1 flag. A 0 flag is assigned when there are no drugs in the assessment period associated with the category. A 1 flag is assigned when there is any drug in the assessment period associated with the category.

At this point in the process, there are quite a few categories. The next step is to reduce and refine the categories. Like the categories themselves, the rules used for reducing and refining the number of categories have been developed independently by each vendor. But the rules are typically "Or," "And," and hierarchical reductions.



TIA GOSS SAWHNEY is the owner of TS Strategic Consulting and can be reached at tia@tsstrategic.com.

An “Or” reduction is used when the new category introduced by the reduction process represents a master category. Therefore, if there is a 1 in any of the sub-categories, then the master category is also assigned a 1. The logic is along the lines of: If $Flag_A=1$ or $Flag_B=1$, then $Flag_C=1$. Master category reductions are used whenever several categories are not deemed sufficiently unique. This might be the case, for example, for multiple categories of drugs related to minor skin conditions.

Other categories introduced by the reduction process represent co-morbidities through interactions. The master category is 1 only if two or more subcategories are 1. This is an “And” reduction. The logic is along the lines of: If $Flag_A=1$ and $Flag_B=1$, then $Flag_C=1$. This type of reduction is used whenever the presence of two conditions is expected to have a larger impact on future health costs than the sum of the impacts normally associated with each condition: $Impact_A + Impact_B < Impact_C$. This may be the case for diabetes and heart disease, which are both of concern individually but when conjoined indicate someone whose illness is greater than the sum of each individual condition. Sometimes this process increases rather than reduces the number of flags, as $Flag_A$ and $Flag_B$ aren’t removed when $Flag_C$ is introduced.

The final type of reduction is a hierarchical reduction. Two or more initial categories are arranged in a hierarchy. Only the flag for the category highest in the hierarchy persists. Therefore, if $Flag_{dominant}=1$, then $Flag_{subordinate}=1$ becomes $Flag_{subordinate}=0$. This method is used when it’s important to identify the most severe category of several related categories.

The result of the reduction process is a series of 0/1 flags related to the master categories. 0/1 flags representing age and gender are then appended to the drug history flags.

Based on the vendor’s statistical studies (multivariate regression models), the vendor assigns an impact value to each po-

tential 0/1 flag, including age and gender. Although users may modify the impact value based on their own statistical studies, the base values are preloaded by each vendor. The impact can be measured either in absolute dollars or in a contribution to an overall risk score. From here on, I will be referring only to risk score.

A risk score, which is the future health care costs of the individual as compared with an average individual, is calculated as the total of each possible impact value, multiplied by the associated 0/1 flags.

$$\text{RiskScore} = \text{Flag}_C \times \text{Impact}_C + \text{Flag}_D \times \text{Impact}_D + \text{Flag}_E \times \text{Impact}_E \dots$$

Where each Flag is either 0 or 1.

This score may subsequently be normalized to the average risk score for that age and gender.

Evaluating the Tools

The beauty of the algorithm just described is that it’s easy to describe. It’s also easy to program, relying on relatively simple rules and tables. And it has proven statistical value. The algorithm does a much better job of predicting future health costs than traditional methods that relied solely on age and sex.

The relative rankings of individual risks, however, often don’t hold up upon examination. This can be traced directly back to the algorithm. While easy to describe and program, the algorithm doesn’t reflect what an underwriter or medical director intuitively knows about assessing future health care risk.

Even though these tools are generally used to evaluate groups of people, not individuals, it’s important that individual scores withstand scrutiny. Better quality individual scores produce better quality group scores, particularly for smaller group sizes, where we can’t assume that scoring errors will be offsetting.

Here are some hypothetical examples where risk scores don’t match medical intuition, based on prescription histories that I’ve reviewed and on my knowledge of the underlying algorithms.

Individual A has a clean drug history except for a single day, several months ago, when he filled three prescriptions: an antibiotic, a decongestant, and a sympathomimetic (a medication that opens airways, often prescribed for asthma). It’s relatively safe to assume that Individual A had the flu.

Individual A, however, gets the same risk assessment as **Individual B**, who fills a sympathomimetic prescription every month and, in addition, routinely fills prescriptions for other drugs commonly associated with asthma.

About a year ago, **Individual C**, a 30-year-old female, filled a couple of prescriptions for an anti-vomiting drug. She filled these prescriptions along with prescriptions for prenatal vitamins and iron, prescriptions that she continued until four months ago. It’s safe to assume that the anti-vomiting drugs were associated with a pregnancy that has ended.

But because the anti-vomiting drug is often given in association with chemotherapy, **Individual C**’s risk score will be boosted (above the score produced by her age and gender) by the same amount as the risk score of **Individual D**, a 55-year-old male, who for the last three months has been taking the drug along with pain medicines and prescription vitamins. **Individual D** is likely to have cancer or a serious digestive disorder.

Finally, **Individual E**, a 20-year-old who filled a five-day prescription for the same drug several months ago and otherwise has no prescription history, will get the same risk score boost. **Individual E** is likely to have had food poisoning that poses no ongoing risk.

Individual F’s only history is for a five-day prescription for pain medicine more than six months ago. She has no boost in her risk score and therefore receives the minimum score for her age and gender.

Individual G, however, has been filling a variety of pain medicines on a regular basis for the past six months. She also receives the same minimum score for her age and gender, even though she clearly has an ongoing health issue.

Finally, **Individual H**, who took a variety of pain medicines until nine months ago and hasn't had a prescription since, also receives the same minimum score.

Individual I has taken the same dose of an antidepressant for the past year and receives a modest boost to her score.

Individual J has taken four different antidepressants in the past six months, including multiple dosage levels for each antidepressant and receives the same boost to her risk score as **Individual I**.

Individuals K and **L** get the same boost in their score even though **Individual K** takes a steady low dose of a high blood pressure drug and **Individual L** filled one 30-day prescription for high blood pressure 12 months ago and another 30-day prescription three months ago. There's every reason to believe that **Individual L** has high blood pressure but is noncompliant.

The risk scores produced by these drug history patterns can be easily explained by the failure of the algorithms to incorporate concepts easily understood by underwriters and medical directors. Concepts not well incorporated into the current algorithms include:

1. Quantity matters—Drugs, often the same drugs, are prescribed for both acute and chronic conditions. Antibiotics and pain medicines are the classic examples. A few days using these drugs represents a fundamentally different risk than several months. We see this with Individuals A and B, Individuals D and E, and Individuals F and G. Quantity, however, has no impact in the current algorithms—any amount of a drug associated with a category raises the flag for that category to 1.

2. Continuity matters—If a drug treats a chronic condition, continuous use is better than episodic use. We see this with Individuals K and L. Since long-term control of blood pressure is important, individuals who take medicine episodically (probably corresponding to their most recent doctor visits) are worse risks than those who

take it steadily. Continuity, however, has no impact in the current algorithms—any amount of drug, continuous or episodic, associated with a category raises the flag for that category to 1.

3. Recent use matters—Recent use of drugs has a bigger impact on future costs than drug use that is clearly in the past. We see this with Individuals G and H. Recent use, however, has no impact in the current algorithms—any drug associated with a category raises the flag for that category to 1, irrespective of when the drug was filled as long as the date was within the assessment period (typically one year).

4. Multiple (concurrent or sequential) drugs within a category matter—A person with stable use of a chronic drug presents an entirely different risk profile from someone whose treatment has not been stabilized. There will be costs associated with achieving stability. This applies to every category of chronic drugs. Furthermore, multiple drugs within a category, either concurrently or sequentially, can be indicative of particularly complex or drug-resistant conditions. We see this with Individuals G, I, and J. Multiple drugs within a single category, however, have no impact in the current algorithms.

5. Concurrent drug categories matter—Current drug algorithms acknowledge the impact of concurrent categories of drugs when they build master categories using "And" logic. But they don't take this logic far enough. First, they look only at concurrent use and the co-morbidities of a limited number of major chronic conditions. They don't look at all important co-morbidities, nor do they look at concurrent drug usage that might actually represent a risk reduction. Individual C is an example of a risk reduction associated with concurrent drugs. Finally, because the models aren't time-sensitive, their embedded definition of concurrent use is inadequate. The actual fill dates associated with concurrent categories may be for prescriptions filled 12 months apart (or more if the assessment

period is longer). This is significantly different from days apart.

6. Age/gender and drug categories are very likely not independent variables—By assigning separate flags to drug categories and to age/gender, the current algorithms assume that they drive future cost independent of each other. Yet, the specific age/gender of a person can sometimes profoundly influence the interpretation of a drug history and the future risk of the individual. We see this with respect to Individuals C and D.

7. Innocuous is not always innocuous—There are many categories of drugs that are dominated by innocuous uses that have no real impact on a person's future risk. These categories include common antibiotics, painkillers, and cough syrups. But while the current algorithms don't assign any value to flags for these categories, there are situations where these drugs matter a lot, particularly when they are being used in regular episodes, over several months, or are currently in use. We see this with respect to Individuals F and G.

8. "Buzz" matters—An excess number of drug category flags indicate someone who is fundamentally unhealthy, a hypochondriac, and/or has a doctor who over-prescribes. None bodes well for future costs. Yet because the current models assign a value to each master category independent of all other categories, the impact of the categories is merely additive except when there is an interaction due to co-morbidities. There's no additional cost boost associated with a large number of categories.

9. Differentiation needs to be made between mean and variance—Insurance companies are in the risk business. They need to differentiate between an individual likely to have \$1,000 of excess claims next year because of a stable (near certain) drug treatment and an individual with an extra 1 percent probability of a \$100,000 risk. Yet because the current models assess only the expected future costs of the individual (mean) and not the risk associated of a

large claim (variance), the two individuals receive the same risk score.

10. The intended purpose matters—An insurer has multiple reasons for wanting to assess risk. Assessment tools predict an individual's future risk, but when they are used for physician compensation, they also can influence physician behavior and the future risk. Lack of continuity, multiple drugs, and/or buzz can bode poorly for the future risk of an individual. But if such factors are reflected in physician compensation, it may be tantamount to rewarding poor medical practice. There may be a legitimate need for different models for different risk-assessment purposes.

Improving the Tools

Based on the above discussion, underwriters, medical directors, actuaries, and even statisticians should agree that the current algorithms are suboptimal. But changing them will be difficult.

Making substantial improvements will destroy their simple elegance. Statisticians value simplicity, and insurance executives (often actuaries) who buy risk assessment tools value simplicity even more. The new algorithms will be complex.

Ultimately, complexity will prevail because it will produce better predictive value. But in order to justify it, the marginal impact has to be proven. Underwriters and medical directors need to use their medical wisdom to give statisticians precise instructions as to the significance of certain drug consumption patterns and see if medical wisdom can be proven statistically.

Both sides will benefit from the interaction. Much of what underwriters understand to be fact today has, in fact, never been statistically quantified. There will undoubtedly be surprises when their wisdom is tested statistically. Certain patterns will turn out to be more or less predictive than underwriters currently presume. Further,

statisticians may show underwriters predictive patterns they previously knew nothing about. A close working relationship will not only lead to better risk assessment tools but will also provide essential knowledge to those responsible for manually underwriting new business.

Risk assessment tool vendors respond to the demands of their customers. Very often, actuaries make or are involved in the purchase of these tools. Isn't it time that we demand tools that hold up better to underwriting and medical scrutiny? Isn't it time for various disciplines to work together and learn from one another? ●

This article is solely the opinion of its author. It does not express the official policy of the American Academy of Actuaries; nor does it necessarily reflect the opinions of the Academy's individual officers, members, or staff.



Do you want a trusted partner to help navigate your actuarial career path?

 **JACOBSON**

Partner with a company that considers your success its top priority. The Jacobson Group has been committed to helping actuaries achieve success throughout all stages of their careers and to assisting employers with their actuarial talent needs for nearly 40 years.

(800) 466-1578 ● WWW.JACOBSONONLINE.COM