Methodological Framework for Propensity Adjusted Benchmarks for Pharmacy Key Performance Indicators

B. Hunter, MS[1]; K. Brown-Gentry, MS[1]; J. Scripture, MS, PhD[1]; D. Eckwright, PharmD[1]; K. Prasla, PharmD, PhD[1]. [1]Prime Therapeutics LLC, Eagan, MN, United States.

Background

- Benchmarks are used to evaluate the performance of drug management programs and are often stratified by line of business without further adjustment. However, subpopulations within line of business can have varying risk profiles associated with population characteristics (e.g., comorbid burden, demographics) that create biased key performance indicator (KPI) comparisons.[1,2]
- Measures of prescription drug utilization have been cited as limited by the lack of risk adjustment for population differences in demographics and health status.[3] Risk adjustment to KPIs reduce noise in the data allowing organizations to monitor progress, make data driven decisions, and measure effectiveness of strategy with reduced confounding influences.[2]
- Many industries (e.g., health care, insurance, finance) are already controlling for these extraneous influences by applying risk adjustments when comparing a population of interest to a benchmark on KPIs. This method could help pharmacy organizations provide more accurate insights when comparing different populations.

Objective

To create a methodological framework for using propensity weights to adjust for varying risk profiles when comparing an index group to a benchmark group on pharmacy KPIs

Methods

Sample

- Commercially insured sample enrolled in a medical drug management program that sought to reduce medical drug spend through a set of services (e.g., fee schedule, prior authorization, step therapy).
- Members had to be continuously enrolled for the entire measurement period (Jan. 1, 2022-Dec. 31, 2022).
- Average age of the analyzed sample was 35 (SD = 20).
- Two customers were analyzed as the index client.
  - Customer A was a small customer (n ~= 5,000) — the available sample for the benchmark group was ~8,000,000.
  - Customer B was a large customer (n ~= 3,000,000) — the available sample for the benchmark group was ~6,000,000.

Propensity Weights

- Propensity weights, which are the probability that a particular individual matches to a population of interest based on a set of covariates, were created to identify a matched benchmark for each customer.
- Covariates were chosen based on the following criteria:
  - Had to significantly predict group membership
  - Had to have a theoretical relationship to the dependent measure
  - Had to have overlap between the index group and benchmark group
- Covariates included in the propensity weight were:
  - Age
  - Therapy count — count of conditions for which the member received medically administered drugs during the measurement period
  - Therapy risk — risk weight for each therapy category based on cost of treatment (determined by allowed amount)
  - Indicator for new start to a particular therapy
  - Claim count
- Covariates were entered into a logistic regression model predicting index group membership.
- Weights for the index group were recoded to 1 so that only the benchmark outcomes were weighted.
- Balance achieved by the propensity weight was assessed by examining group differences as the standardized difference after applying the weight.

Outcomes

- Dependent measure — per member per month (PMPM) cost for medical drugs was the primary outcome.
- PMPM medical drug cost was also stratified by the top five therapy conditions and medications.
- Unweighted and weighted means were calculated for benchmark group for comparison to the index client means.

Table 1

Propensity balance for covariates included in propensity weight calculation after weighting benchmark to match Customer A

	Index – Customer A	Weighted	
		Benchmark	Cohen's d
Age	42.96	42.96	0.01
Therapy Risk Score	0.03	0.03	0.00
Therapy Count	0.16	0.15	0.04
New Start Drug Count	0.12	0.12	0.07
Claims Count	0.15	0.15	0.02

Notes: Values represent the sample mean for each covariate included in the propensity score. Cohen's d is the standardized difference between the two group means calculated as (Index Mean – Benchmark Mean)/pooled standard deviation. Values < 0.10 indicate that good balance has been achieved by the propensity weight.

Table 2

Propensity balance for covariates included in the propensity weight calculation after weighting benchmark to match Customer B

	Index – Customer B	Weighted	
		Benchmark	Cohen's d
Age	38.16	37.94	0.02
Therapy Risk Score	0.04	0.03	0.02
Therapy Count	0.16	0.15	0.02
New Start Drug Count	0.11	0.10	0.01
Claims Count	0.15	0.14	0.03

Notes: Values represent the sample mean for each covariate included in the propensity score. Cohen's d is the standardized difference between the two group means calculated as (Index Mean – Benchmark Mean)/pooled standard deviation. Values < 0.10 indicate that good balance has been achieved by the propensity weight.

Table 3

Comparison of unweighted and weighted benchmark comparisons for Customer A

PMPM	Unweighted		Weighted
	Index – Customer A	Benchmark	Benchmark
Allowed Amount PMPM	$18.60	$39.67	$23.03
Top 5 Therapies			
Oncology	$5.85	$14.22	$4.64
Crohn's & Ulcerative Colitis	$1.82	$4.75	$6.87
Multiple Sclerosis	$1.68	$3.96	$1.95
Immune Globulin	$3.66	$2.59	$2.21
Colony Stimulating Factor	$0.23	$1.50	$0.65
Top 5 Drugs			
Keytruda	$3.64	$3.27	$0.95
Ocrevus	$1.68	$3.00	$1.16
Opdivo	$0.00	$1.22	$0.40
Eylea	$0.30	$0.45	$0.39
Entyvio	$0.72	$2.21	$3.38

Notes: PMPM values were calculated by summing the total allowed amount for medical drugs and dividing by the monthly membership. Claims were limited to the therapy or drug of interest when calculating the PMPM values for the therapy and drug stratifications.

Table 4

Comparison of unweighted and weighted benchmark comparisons for Customer B

PMPM	Unweighted		Weighted
	Index – Customer B	Benchmark	Benchmark
Allowed Amount PMPM	$46.46	$37.33	$47.33
Top 5 Therapies			
Oncology	$17.09	$13.23	$17.97
Crohn's & Ulcerative Colitis	$4.34	$4.90	$5.28
Multiple Sclerosis	$4.65	$3.72	$4.38
Immune Globulin	$3.19	$2.39	$2.96
Colony Stimulating Factor	$1.76	$1.41	$1.97
Top 5 Drugs			
Keytruda	$3.39	$3.22	$4.54
Ocrevus	$3.49	$2.84	$3.26
Opdivo	$1.23	$1.22	$1.65
Eylea	$0.54	$0.42	$0.54
Entyvio	$1.83	$2.34	$2.52

Notes: PMPM values were calculated by summing the total allowed amount for medical drugs and dividing by the monthly membership. Claims were limited to the therapy or drug of interest when calculating the PMPM values for the therapy and drug stratifications.

Results

- Table 1 displays the mean values for each covariate included in the propensity weight for Customer A compared to the weighted benchmark group. After applying the weights to the benchmark sample, all differences in covariates between the index and benchmark were reduced to Cohen's d < 0.10. Cohen's d is the absolute standardized difference between the index and weighted benchmark, and values < 0.10 provide support that good balance was achieved for a given covariate.
- Table 2, like Table 1, displays the mean value for each covariate included in the propensity weight for Customer B compared to the weighted benchmark group. Again, all values for Cohen's d were < 0.10.
- Table 3 displays the PMPM for Customer A as well as the unweighted and weighted PMPM for the benchmark. Prior to weighting, one might conclude that Customer A is performing much better than the benchmark ($18.60 vs $39.67, respectively). However, after weighting the benchmark mean to create a matched sample for the index group, we see that while Customer A is doing better than benchmark, the magnitude is much smaller than it appeared based on the unweighted benchmark PMPM ($18.60 vs. $23.03, respectively).
  - Additionally, Table 3 provides PMPM allowed amounts stratified by top therapy categories and drugs. One of the key drivers, for example, contributing to the large magnitude of difference between Customer A and the unweighted benchmark was oncology spend. By weighting the benchmark to match Customer A, we see that the oncology spend is comparable between the two groups now that the index customer is being compared to a benchmark sample with a similar oncology burden. This was a key factor in the smaller magnitude of difference between Customer A and the weighted benchmark regarding the overall PMPM allowed amount.
- Table 4 displays the PMPM for Customer B as well as the unweighted and weighted PMPM for the benchmark. Prior to weighting, it appears that Customer B is not doing as well as the benchmark ($46.46 vs. $37.33, respectively). However, after weighting the benchmark to match the index client, we see that Customer B is performing approximately the same, if not slightly better, than the matched benchmark ($46.46 vs. $47.33, respectively).
- Table 4 also includes the PMPM allowed amounts stratified by top therapy categories and drugs. Again, one of the key drivers contributing to the difference between Customer B and the unweighted benchmark was oncology spend. In this scenario, when the benchmark is weighted to match Customer B, we see that oncology spend is comparable between the two groups, which contributed to the near equivalence of the overall PMPM allowed amount between Customer B and the weighted benchmark. Weighting the benchmark will not always result in decreasing the gap between the two groups, rather it may increase the gap in some cases. Crohn's and ulcerative colitis spend demonstrates how controlling for risk profiles between the index and benchmark by using a propensity weight can reveal a greater magnitude of difference between groups.

Limitations

- Important potential covariates for inclusion in the propensity score may have been excluded due to data limitations. For example, sex and data pertaining to social determinants of health were not available for inclusion. It is possible that unobserved variables could still be biasing the benchmark comparison.
- The analysis was conducted on a commercially insured population and results may not generalize to other lines of business.
- This analysis is limited to medical claims data and does not capture any drugs adjudicated on the members' pharmacy benefits.

Conclusion

- When comparing an index group to a benchmark with varying risk profiles, KPIs may not reflect the index group's performance, but instead reflect differences in population composition, such as a greater disease burden.
- Accounting for confounding differences through a propensity weight can control for varying risks across groups and distill a more meaningful comparison by comparing the index group to a benchmark with a similar risk profile.

References

1. Schweizer ML, Braun BI, Milstone AM. Research methods in healthcare epidemiology and antimicrobial stewardship–quasi–experimental designs. Infect Control Hosp Epidemiol. 2016;37(10):1135–1140. doi:10.1017/ice.2016.117
2. Risk adjustment. Centers for Medicare & Medicaid Services. Accessed August 15, 2024. https://www.cms.gov/priorities/innovation/key-concepts/risk-adjustment#:~:text=Risk%20adjustment%20is%20used%20to,the%20patient's%20specific%20health%20needs
3. Chawla AJ, Hatzmann MR, Long SR. Developing performance measures for prescription drug management. Health Care Financ Rev. 2001;22(3):71–84.