Using Between-Patient Comparisons to Quantify Outcomes

A Brief Guide

J. Brian Cassel, PhD, Kathleen Kerr
December 2021

This guide summarizes several of the key decision points and choices for plans that wish to explore approaches for assessing financial and other outcomes for their palliative care programs and quantify program impact by comparing outcomes for members who received palliative care to outcomes for members who did not receive palliative care.

Decision #1: Whether and when to use a between-persons approach.

Between-persons approaches (i.e., program recipients compared to others) are more difficult and time-consuming than within-person (i.e., pre vs post) approaches. Thus, the first decision is whether and when to use one approach or the other.

Much can be learned about program implementation by routinely examining volumes, recipient characteristics, provider characteristics, costs, and duration of program enrollment among those enrolled. In that context, all programs would benefit from conducting pre vs. post outcome analyses early in implementation to see if intermediate outcomes are meeting expectations or are at least trending in the desired direction. For example, palliative care programs may be expected to decrease emergency department visits and symptom-focused hospitalizations, and those can be assessed using a within-persons approach that assesses event frequency several months prior through several months after program enrollment.

Once program implementation has matured – when volumes and length of service reach levels that are desired, and pre- vs. post-treatment outcomes have been evaluated – it may be useful to quantify outcomes using more rigorous methods that feature a well-constructed comparison group. For palliative care programs, assessing fiscal impact using a between-persons approach can yield more reliable results than a within-persons approach, given the increase in total costs of care in the final year of life that is associated with usual care, and the reduction in costs in the final months of life that has been reliably demonstrated in the palliative care literature. For example, if a within-persons analysis showed that among palliative care program enrollees costs the final year of life were no different than costs in the year before death, it may appear that the palliative care program had a neutral financial outcome. However, a between-persons analysis of the same data set might show that holding costs steady in the final year of life is indicative of significant cost savings. The opposite can also be true, where reductions in costs of care seen in the months following enrollment might be found to be non-significant when tested with a between-persons analysis.

Conducting between-persons analyses is considerably more difficult than pre-post analyses, so one must balance the importance of the information that will be garnered with the time and resources that will be expended. Questions to consider include:

- <u>Value</u>: What will we learn from more rigorous analyses? What will we do with the information?

- <u>Timing</u>: Is the program implementation sufficiently mature, and are we hitting targets for process and intermediate outcomes?

- <u>Resources</u>: Do we have sufficient data and analytic personnel to do this?

Decision #2: To whom should recipients be compared?

One of the central challenges for between-persons comparisons is ensuring that the enrolled and comparison groups are adequately similar. Published studies of home-based palliative care programs in the US have compared recipients to either other decedents who were not enrolled in the program, or other beneficiaries who were eligible for the program but who did not enroll.

<u>Decedents approach</u>. Several studies have focused analyses on healthcare utilization and costs in the final months of life, comparing deceased palliative care recipients to similar decedents who did not receive palliative care. Outcomes have included total costs of care, hospitalizations, hospice enrollment, and hospice duration. This approach requires reliable data on date of death for health plan beneficiaries. For Medicare and Medicaid lines of business, this may be available from CMS. For commercial insurance, external death data sources such as state departments of health and/or the national Social Security death data would be needed.

<u>Eligible beneficiaries approach</u>. Several studies – including both randomized controlled trials and retrospective, observational studies – compare palliative care recipients to similar beneficiaries who were eligible for, but did not receive, the service. In such studies, the two groups can be compared from the point of eligibility forward, not necessarily limited to those who are known to be deceased. In some studies, the comparison group is limited to those who were fully screened or even offered the service. For example, the Mayo clinic study (Chen 2018) identified dozens of patients who fully met service criteria, but who could not be served due to limited clinical capacity.

Decision #3: Analyze by subgroups, or leave the whole population intact?

Because beneficiaries with different insurance types – Medicare Advantage, commercial, and Medicaid – may differ greatly in terms of their ages, primary serious illnesses, co-morbidities, socio-economic resources, social risks, etc., we recommend that analyses of palliative care programs be stratified by line of business. Thus, Medicaid enrollees, for example, would be matched or compared only to other Medicaid enrollees, and results would be reported separately for each line of business. This is especially important when the health plan is responsible for the cost of a modality in one line of business, but not another – for example, for most Medicare Advantage plans, the cost of hospice is not borne by the MA plan. Health plans with palliative care programs across multiple lines of business are advised to analyze each line of business separately.

Other variables – geographical area, primary disease, demographics – are usually incorporated into matching, but results are not reported separately. One study (Cassel 2016) did stratify by disease type

(cancer, CHF, COPD, or dementia) for matching and reporting of some results, but the disease type was not a significant predictor of outcomes, and the four disease groups were combined for reporting of some findings.

Decision #4: How exactly to create the comparison group?

Aside from a true experiment (Brumley 2007), most studies of home-based palliative care use propensity-based matching or weighting to construct a comparison group. This approach is described in detail in several palliative care studies (Garrido, 2004; May 2018; Cassel 2016). The goal is to use beneficiaries' baseline characteristics to create well-matched samples of recipients and non-recipients.

Propensity scores reduce multiple variables into a single score representing the likelihood that an individual would have received treatment. The propensity for treatment is derived from a logistic regression analysis where treatment receipt (yes or no) is the dependent variable, and predictors are drawn from data known about the beneficiaries at baseline. Using propensity scores, recipients and non-recipients are matched according to baseline similarity (before treatment was offered). Typical variables from claims data that are used for matching include demographics, primary disease, co-morbidities, baseline hospital use, geographic area, costs of health care, and the like.

The use of propensity scores in health services research is quite common. While true randomization will usually create groups that are equal in both measured and unmeasured variables, propensity scores only allow you to balance on measured variables. As such, studies that rely on propensity scores are not an equal substitute for true randomization. In fact, it may take numerous repetitions of the steps – recategorizing or dropping predictor variables, checking balance of groups, assessing cases dropped – to achieve adequate (though imperfect) balance between recipients and non-recipients.

Note that while some published studies of home-based palliative care do not use propensity scores for construction of the comparison group, we strongly recommend that a rigorous method such as propensity scores be used, to guard against accidently matching complex, seriously ill beneficiaries who used palliative care with others who were not as complex, seriously ill, or at risk. The propensity score approach doesn't just make the two groups similar, it also matches each program recipient to the non-recipient(s) that are most similar.

Decision #5: Matching vs. weighting

Once propensity scores are created, they can be used in three slightly different ways to construct a comparison group. One option used in several studies is to allow multiple non-recipients to be matched to each recipient. For example, in the Sharp study, most recipients were matched to three non-recipients: 368 palliative care recipients were matched to 1,075 non-recipients. Another option is to match 1:1, which has been done in the study of the Sutter AIM program (Ruiz 2017). A third option is to use propensity weighting, which retains more non-recipient persons, and just assigns them a lower weight in the analyses. The weighting approach was used in a recent meta-analysis of inpatient palliative care (May 2018). For guidance on the complex question of which of these three approaches to use, see Garrido et al., 2014.

Decision #6: What time periods to compare?

It is critically important to know when exactly the recipients were enrolled in palliative care and to distinguish the pre-intervention period from the post-intervention period, as in any evaluation of treatment effects.

Defining enrollment or pseudo-enrollment date. Once enrollment dates are known for recipients, a pseudo-enrollment date can be created for the non-recipients. In decedent studies this is typically done by looking backwards from death for the same number of months for the non-recipients as for each recipient to whom they were matched. This is relatively non-controversial. For studies that include survivors, setting a pseudo-enrollment date for non-recipients is more challenging. If there is a specific event that occurs that determines eligibility – staff referral, hospitalization, or date of screening for example – the timing of that event may be used to set the pseudo-enrollment date. The literature offers little guidance on how to define a pseudo-enrollment date in the absence of a specific event that determines eligibility.

Handling dis-enrollment. Inevitably, program enrollment will end for some home-based palliative care recipients prior to death. Some may transition to hospice care. Some may initiate disenrollment, indicating they no longer want or need the service. Some may no longer meet eligibility criteria – for example physically rebounding and entering a phase when ambulatory visits to a palliative care provider are now possible, instead of receiving home-based care. In most home-based palliative care studies, the outcomes continue to be measured after recipients end services, the assumption being that the effect of the intervention is assumed to continue even after disenrollment.

Use of per-month measures. Because there is likely to be significant variation in length of enrollment in palliative care services, some studies measure hospitalizations and cost of care in units per month. This is especially useful in evaluating value-based contracts where providers are reimbursed with per-enrolled-member-per-month payments.

Decision #7: Quantifying differences – what didn't happen?

Once the treatment and comparison groups are constructed and the time frames are determined, one can compare outcomes for the recipients to controls, and use appropriate statistical methods (e.g., 95% confidence intervals) to evaluate statistical significance. It is important to separate the time before treatment from the time in treatment. In some studies, using eligible non-recipients as controls, only the post-treatment time was compared (Yosick 2019). In one study, a difference-in-differences approach was used in which the difference between the baseline year and the post-eligibility year were compared for the treatment and control groups (Chen 2018). This is a strong approach for quantifying outcomes for the two groups by combining elements of the pre- vs. post and between-persons analyses. The CMS evaluation of the Medicare Care Choices Model (CMS 2020) used a difference-in-differences analysis of decedents as well, and implementation was randomized at the hospice level.

Similar survival rates? Ideally, the palliative care recipients and the non-recipients will have equal survival. However, in most real-world studies, palliative care recipients are more likely to die than comparison patients. That is one reason many studies use the decedent approach, and the time from enrollment to death is matched for the non-recipients. For approaches using eligible non-recipients as controls, it is possible that the comparison and treatment groups will have differences in complexity and

seriousness of illness, and thus in survival as well. There are two potential problems arising from this. One is that you will be comparing a sicker and more costly group (palliative recipients) to a less sick, less costly group (non-recipients). The difference-in-differences approach may help to adjust for this. A very different issue is called a "competing risks" problem. People who are dead are no longer incurring costs, and if the timeframe of analysis is not adjusted for decedents, then they may have lower risk for all forms of utilization. This was addressed statistically in one study of inpatient palliative care and effects on hospital re-admissions (May 2019) but has not been explicitly addressed in home-based palliative care studies that use eligible non-participants (rather than decedents) for controls.

Summary of decisions to make, and recommendations

Decision	Recommendation
Do between-persons analyses? When?	Yes, do between-persons analyses when 1) program implementation is mature and on-track; 2) you will derive some clear value from the results; 3) you have analytic resources available including high-quality data and analytic bandwidth.
Compared to whom?	Use a decedent cohort approach if you are most interested in end-of-life utilization and costs, and if you have trustworthy data on dates of death. Use eligible non-recipients if there is a clear eligibility event such as clinical screening or referral. Either may be justifiable depending on the program type and the question you are trying to answer.
Analyze by subgroups?	Health plans with multiple lines of business should analyze each line of business separately.
How to create the comparison group?	If an experiment (randomly assigning beneficiaries, providers, or regions to service or eligibility) is not possible, use propensity score methods to reduce unwanted variation between groups and match recipients to similar non-recipients.
Match? Or Weight?	Either approach is justifiable.
What time periods?	Separate pre-treatment from treatment periods. Match treatment time period of non-recipients with those of recipients. Consider measuring outcomes from enrollment until death even if dis-enrollment occurred. Consider measuring outcomes by month.
How to quantify what didn't happen?	Compare recipients and non-recipients on outcomes of interest (e.g., hospitalizations, hospice enrollment, hospice duration, total costs of care). Evaluate length of survival in both groups and consider whether "competing risks" could be an issue for studies using eligible non-recipients.

References

Paper	Program	Methods
[Brumley JAGS 2007](#)	Kaiser Permanente (multiple sites)	RCT – randomized at patient level.
[C. Kerr JPM 2014 (cost)](#) [C. Kerr JPSM 2014 (other outcomes)](#)	Home Connections (Buffalo)	Decedent cohort, propensity score matching, ~3:1.
[Lustbader JPM 2016](#)	Prohealth (NYC region)	Decedent cohort, not matched.
[Cassel JAGS 2016](#)	Sharp "Transitions" (San Diego)	Decedent cohort, propensity score matching within disease group, 3:1.
[Ruiz Innov Aging 2017](#) [Ruiz Health Aff 2017.](#)	Sutter AIM (Northern Calif)	Decedent cohort, county-level similarity, propensity matched, 1:1.
[Chen JPSM 2018](#) (cost). [Chen JPM 2015](#) (model, other outcomes)	Mayo Clinic (Minnesota)	Eligibles, propensity-matched, 2:1.
[Yosick JPM 2019](#)	Trinity / Turnkey (Columbus OH)	Eligibles, propensity-score adjusted, ~3:1.
[CMS 2020](#)	MCCM 3rd year evaluation (multiple sites)	Randomized at hospice level; decedent cohort; difference-in-difference analyses.
[Gordon Le JPM 2021](#)	Healthnet California	Eligibles, not matched.

Other references that detail statistical methods for observational studies of palliative care programs

Garrido MM, Kelley AS, Paris J, et al. Methods for Constructing and Assessing Propensity Scores. Health Serv Res. 2014;49(5):1701-20.

May P, Garrido MM, Cassel JB, et al. Prospective Cohort Study of Hospital Palliative Care Teams for Inpatients with Advanced Cancer: Earlier Consultation Is Associated With Larger Cost-Saving Effect. J Clin Oncol. 2015 Sep 1;33(25):2745-52. PMID: 26056178

May P, Normand C, Cassel JB, et al. Economics of palliative care for hospitalized adults: a meta-analysis. JAMA Intern Med. 2018 Jun 1;178(6):820-829. Doi: 10.1001/jamainternmed.2018.0750. PMID: 29710177

May P, Garrido MM, Del Fabbro E, et al. Evaluating Hospital Readmissions for Persons with Serious and Complex Illness: A Competing Risks Approach. Med Care Res Rev. 2019. Doi: 10.1177/1077558718823919. PMID: 30658539