Appendix 1 Sampling strategy

1.1 Site selection

Our identification of hospitals for inclusion in this study is detailed §1.1.1-1.1.3 and summarised in Figure S1.

1.1.1 Hospital Inpatient Enquiry (HIPE) database

Our main data source is HIPE, which records every admission to a public hospital in Ireland. According to the Healthcare Pricing Office (HPO), which operates HIPE, the database has run since 1971 and there are approximately 1.7 million HIPE records created annually.(https://www.hiqa.ie/areas-we-work/health-information/data-collections/hospital-patient-enquiry-hipe)

There are 56 public hospitals listed in the HIPE data dictionary (https://www.hpo.ie/hipe/hipe_data_dictionary/HIPE_Data_Dictionary_2022_V14.0.pdf). To comply with conditions of data use we do not identify the hospitals at any point in our paper, and we are not permitted to do so; we also limit descriptive data on included hospital sites for the same reason.

Our ultimate aim was to analyse adult inpatient stays following emergency admission; in this context we removed 22 hospitals from the list ex ante as irrelevant either due to population (e.g., maternity and children's hospitals) or hospital type (e.g. orthopaedic hospitals, sites without an emergency department that primarily provide secondary care); there were therefore 34 potentially relevant hospitals in HIPE.

In dialogue with data controllers prior to requesting data, study investigators were told that 2009 is the earliest year for which data are complete and reliable. The most recent available year was 2019.

1.1.2Hospital PMS database

For the 34 acute public hospitals in HIPE that are potentially relevant for inpatient palliative care, we compiled our own database of when each PMS was implemented by contacting hospitals and PMS teams directly through publicly available contact information and our own networks. All 34 hospitals had a PMS by the end of 2016 (1).

To be eligible for our study, hospitals had to fulfil the following criteria:

- Responded to our enquiries to confirm time of PMS implementation, or published data from 2006 allowed us to confirm that a PMS already existed at that time (2).
- No PMS prior to 2009. We excluded all sites with the PMS at the start of the dataset since so-called "always-treated" observations may bias results; see Appendix 2 for the methodological rationale.

Six hospitals/teams did not respond to our enquiries. Of the 28 eligible sites for which we had PMS data, 19 had some level of consultant-led PMS activity by 2009 and nine did not.

1.1.3 Preliminary assessment of HIPE data in eligible sites

For the nine eligible sites with sufficient PMS data, we extracted a provisional sample of admissions for inspection. We extracted from HIPE all admissions at these nine sites in the period 2009-2019 that had in their ICD-10 codes EITHER a diagnosis of serious life-limiting disease [cancer, major organ failure, ADRD, Parkinson's disease (3)] AND/OR Z51.5 indicating palliative care involvement (4).

One hospital did not use the Z51.5 code before or after PMS implementation and was excluded since it was impossible to identify treated patients in HIPE. All other hospitals did use the Z51.5 code following implementation and in a negligible fraction of admissions prior to implementation.

Two hospitals had markedly longer LOS to the other eligible hospitals, reflecting an unusually high proportion of pre-arranged (elective) admissions. A further two hospitals had trends in LOS and mortality rates that were both different from the other four hospitals and from each other. These data undermined the common trends assumption required for our analysis and so these hospitals were excluded.

1.2 Analytic sample

There were approximately 100,000 adult admissions to the four eligible sites in the HIPE database for the years 2009-2019.

We extracted those considered potentially eligible for palliative care: present in their recorded ICD-10 codes, either a

© Annals of Palliative Medicine. All rights reserved. https://dx.doi.org/10.21037/apm-23-479

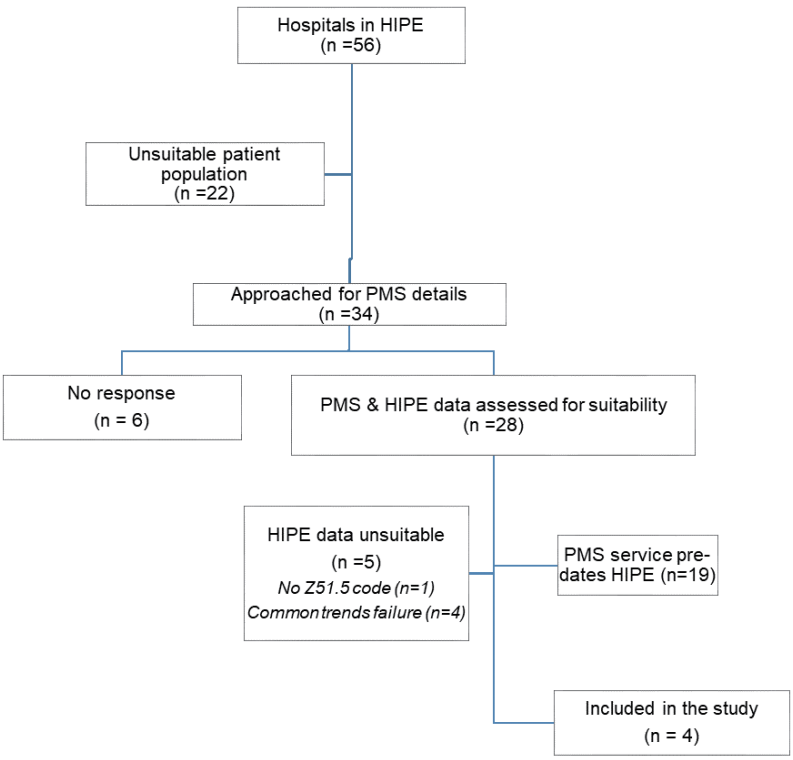


Figure S1 Site selection for the study.

diagnosis of cancer, major organ failure, dementia or Parkinson's disease, AND/OR a Z51.5 code. This was approximately 50,000 unique admissions, approximately 4,500 admissions per year.

For each hospital in the years following PMS implementation, we examined prevalence of a Z51.5 code indicating palliative care interaction. To establish a suitable annual cell size at the hospital level, we identified the year where the number of admissions with a Z51.5 was highest, on an assumption that this reflects reasonable PMS capacity at each site. We denote this value for each hospital, h, as n^h_{max}.

For any given year, y, the hospital-level sample size n^h_y was set equal to n^h_{max}. For each hospital at each year we used propensity score weights using age, sex and diagnoses of serious illness (*Table 1*, main manuscript).

- for every year prior to PMS implementation we retained $n^h_y \left(= n^h_{max}\right)$ observations with the largest weights;
- and for every other post-implementation year we retained all admissions with a Z51.5 code, sample size denoted $n^i_{y,z}$, and rounded out the panel with $[n^i_{max} - n^i_{y,z}]$ observations.

Thus, we created a balanced panel with the same number of observations for each hospital at each year, this included all palliative care interactions recorded in the data, and additional comparison observations most similar those admissions that did involve an interaction. E.g. if for a given hospital, i, $n^i_{max} = 100$ then:

- for every year prior to PMS implementation we retained those 100 observations deemed most likely to have received palliative care had the service been in place;
- and for every other post-implementation year we retained those 100 admissions that included all those with a palliative care interaction plus the additional required observations most likely to have received palliative care *had the service been seen as many patients it did in the year when the largest number of PMS interactions was observed*.

This process identified 7,909 unique admissions for 2009-2019, 719 per year. For the period of observation there were therefore 4,314 unique admissions (per *Table 1* of the main manuscript) (Table S1). This constituted about 5% of all 80,000 admissions at the four sites across the time period.

Distribution of admissions across hospitals is provided in *Table 1* of the main manuscript.

© Annals of Palliative Medicine. All rights reserved.
https://dx.doi.org/10.21037/apm-23-479

Table S1 Number of observations in the sample, all admissions at each site for 2010-2016

Site	1	2	3	4	ALL
Total N in analytic sample=	2,622	876	162	654	4,314
Annual n in analytic sample=	437	146	27	109	719
Total admissions for 2010-2016*					~80,000
What proportion of all admissions were in analytic sample*					~5%

*, data access rules prevent us from publishing identifiable information about any hospital. We therefore restrict these data to pooling the whole sample, and rounding the number of admissions to the nearest 5,000, and proportion of admissions (i.e., 4314/~80,000) to the nearest whole number.

Appendix 2 Analytic framework

2.1 Difference-in-differences: basic set-up

The basic set-up has been in use in health research for well over a century, and was popularised in contemporary economics by Card and Krueger (1993).

For evaluation of a non-randomised intervention (or treatment, policy change, etc.), investigators might compare pre and post outcomes for those who received it, or they might cross-sectionally compare post outcomes for those who received it with post outcomes for those who did not. The before-and-after framework is limited because it cannot control for time trends; the cross-sectional evaluation is limited because of concerns over selection bias.

Difference-in-differences combines these methods, aiming to overcome both bias due to time trends and due to selection. Investigators measure the change in outcome before and after the intervention in a group exposed to that intervention, against the change in outcomes in the same time periods among a group not exposed to that that intervention.

	Pre-intervention	Post-intervention
Exposed group	A (not yet treated)	B (treated)
Unexposed group	C (never treated)	D (never treated)

The intervention effect is thus estimated as [(B-A)-(D-C)] – i.e., we calculate the difference in outcomes over time for the two groups, and the difference between those differences is the treatment effect of interest.

Under certain assumptions, particularly as regards the comparability of groups in characteristics and time trends, this represents a credible causal estimate of intervention effect. Introductory-level materials and more advanced guidance are widely available (5,6).

2.2 Difference-in-differences: multiple time cohorts and time periods

In practice, multiple groups receive an intervention and at different times. Consider Figures S2,S3, images of simulated data sourced from the World Bank blog by Duhat *et al.* (7).

2.1 corresponds to the basic set-up outlined above: Cohort 1 is unexposed, Cohort 2 is exposed in 2005. 2.2 introduces a third group exposed in 2010.

Where Figure S2 provides only one possible comparison, Figure S3 provides three permutations. A lively and voluminous econometrics literature has in the last five years sprung up debating the appropriate approaches to managing multiple groups and time periods (6). In particular, problems emerge when one or more cohorts are treated in both the pre and post period—e.g., Cohort 2 *vs.* Cohort 3 in Figure S3 uses as a comparison Cohort 2, which is 'always treated'. Identified problems are varied, but relate to bias introduced through dynamic and heterogeneous treatment effects over time. Solutions to these problems have also been explored independently by multiple groups (6).

© Annals of Palliative Medicine. All rights reserved.
https://dx.doi.org/10.21037/apm-23-479

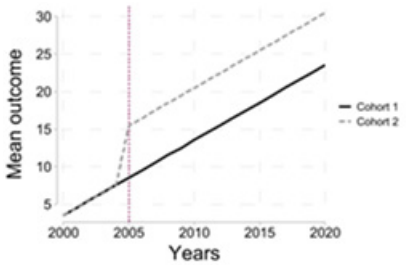


Figure S2 Basic set-up.

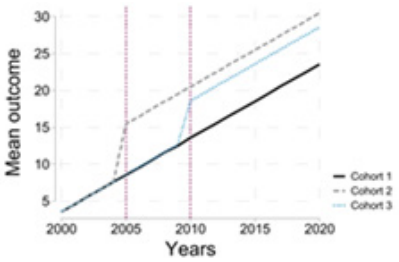


Figure S3 Multiple groups, periods.

2.3 Difference-in-differences: our approach

Our identification of hospital sites and patient observations in the context of available data is detailed in Appendix 1. Data were available from 2009. We excluded at the outset those hospitals that already had a palliative medicine service in 2009, since these would in any head-to-head comparison be always treated, and we additionally excluded hospitals whose outcome distributions contravened comparability assumptions. Our final analytic sample was drawn from four hospitals, whose PMS implementation variously occurred in 2011, 2012, 2013 and 2014. Therefore in primary analysis our time frame was 2010-2015; based on 719 episodes per year (see section 1.2) we had 4,319 episodes organised in 719 six-year cohorts using propensity score matching.

In comparing 2010 and 2011 outcomes we have one treated group and three never-treated hospitals; for subsequent year-on-year analyses to 2015 we have a falling number of never-treated and a growing number of always-treated observations. We excluded all observations from 2016 onwards since at this point the dataset consists only of always-treated observations, leaving us with a six-year (2010-2015) panel.

There are essentially two approaches to managing potential biases arising from including always-treated observations in the panel. First, remove always-treated observations from the analysis so that problematic comparisons do not occur. Second, flexible approaches to estimating treatment effects over time, estimating treatment effects that are weighted according to specific parameters of interest.

We adopt the first of these approaches. Specifically, in primary analysis we retain all observations for all hospitals for all years 2009-2015 and we estimate a standard two-way fixed effects model

$$y_{it} = \beta^{TWFE} I_{it} + \theta_i + \delta_t + X_{it}\beta + \varepsilon_{it}$$

Where *i* denotes one of the 719 cohorts, t denotes year, *y* is the outcome for a given admission, β^{TWFE} is the two-way fixed effects estimator of the average treatment effect on the treated, I is the intervention (did the episode involve timely palliative care?), θ^i is an individual fixed effect, δ^t is a year-specific fixed effect, $X_{it}\beta$ is a matrix of individual-level characteristics that potentially vary from year to year within a cohort (age, sex, diagnoses) and εit is an episode-specific error term.

We then check the robustness of our results to retention of 'always-treated' observations by evaluating each hospital implementations separately in a basic framework; retaining only a never-treated comparison cohort.

Our rationale for this approach is three-fold. First, we consider the risk of heterogeneous treatment effects over time within cohorts to be relatively low. While such dynamic effects bias comparisons of unique individuals, the composition of our dataset combines unique episodes into cohorts. We see little reason to hypothesise that the outcomes of a seriously ill person in a given hospital in year t will materially impact the outcomes of a different (but similar) seriously ill person in that hospital in years $t + \_n$. It may be the case that broader temporal trends exert heterogeneous treatment effects (e.g. as the PMS becomes more established they become more influential in decision-making, but we control for temporal trends in our main analysis and conduct sensitivity analyses with different timeframes imposed). Second, the more technical solutions to this problem require specification and derivation of treatment effect estimates for specific years, sites and sub-groups that are beyond our data. As detailed in Appendix 3, we specify an exposure variable 'timely palliative care' that is fuzzy in definition due to data limitations. In our view there is little point in precision engineering treatment effect estimates of a fuzzily defined treatment; future studies with better data linkage can aspire to more specific estimation of effects under specific conditions.

© Annals of Palliative Medicine. All rights reserved.
https://dx.doi.org/10.21037/apm-23-479

Third, it has been noted that different solutions to treatment effect heterogeneity typically offer similar results, that some solutions are extremely complex in conception and execution, and that TWFE is often more suitable for multiple groups and periods than has been assumed (8). Given the difficulties in conducting and interpreting complex solutions to what may be intractable problems in our data, we prefer via sensitivity analyses multiple analyses of the same more basic set-up as it's preferable to have (relatively) simple results whose interpretation, strengths and weaknesses are (relatively) clear than complex analyses whose meaning is unclear.

Appendix 3 Treatment group definition

3.1 The problem

Evaluation of hospital inpatient palliative care on utilisation must take account of the timing of first palliative care interaction: how many days was it from admission to that interaction?

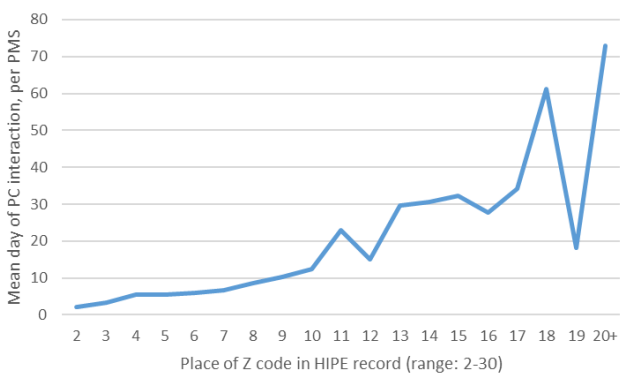
The underlying logic of this is clear and intuitive: utilisation outcomes such as length of stay and costs are accumulated from the moment the admission starts; if the treatment under evaluation is first received when a large proportion of the outcome has already been accumulated, e.g. on day 10 of a 12-day stay, then the treatment cannot be expected to impact outcome. Failure to account for timing biases results to the null. The systematic relationship between timing and effect on utilisation has been demonstrated in both applied and simulated studies previously (9,10).

The practical consequences in research are much less clear. There are no guidelines to define independently what qualifies as "early" or "appropriate" palliative care interaction following hospital admission. Moreover, the date of palliative care interactions are not recorded through HIPE; the only way to identify the precise date of a palliative care interaction is locally through a hospital PMS database. Linking PMS data to HIPE data at each eligible site locally was not feasible for this study.

3.2 Our approach

HIPE records palliative care interactions through the ICD-10 code Z51.5. Presence of this code is not synonymous with a PMS database record but has been shown to capture a large proportion of interactions and has been recommended previously for defining palliative care as a treatment variable in research (4). One key limitation of the Z51.5 code in cross-sectional analysis is selection bias and the ex ante predictability of individual deaths, but our difference-in-differences framework reduces this concern: observations in the comparison group are those who could not receive palliative care because the service did not exist yet, and based on age, sex and diagnostic profile they resemble most closely those who were treated in later years.

We therefore use the Z51.5 code to identify palliative care patients in the post-PMS implementation years at each hospital, and we control for timing of that interaction by controlling for when the code appears in the HIPE record. Each HIPE record allows for up to 30 unique codes and codes are added to records on a rolling basis through the admission. We hypothesised that earlier entry of the Z51.5 code in the record was associated with timing of PMS interaction: the earlier in the HIPE record, the earlier the interaction occurred. We verified this hypothesis by accessing over 2000 PMS records at one hospital and matching to HIPE locally. We identified first PMS interaction in the PMS data and Z code entry in the HIPE data and cross-referenced them. The relationship is unambigious:



© Annals of Palliative Medicine. All rights reserved.
https://dx.doi.org/10.21037/apm-23-479

In our primary analysis we defined timely PMS interaction as occurring in the first five fields of the HIPE record. Per this secondary data analysis, the mean time from admission to interaction in this group was 4.88 days:

```
. su pcday if zcday<=5
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pcday	419	4.883055	8.598329	0	101

This compares to a mean time from admission to interaction in later Z code entries of 18.3 days:

```
. su pcday if 5<zcday
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pcday	1,881	18.33227	37.22207	0	534

3.3 Conclusion and interpretation

Our treatment variable in the primary analysis is, did the admission involve a PMS interaction in a timely way following admission? We define this timeliness in terms of, did the ICD-10 code for palliative care appear in the first five diagnostic items of the HIPE record, which is a group that on average sees a PMS within five days of admission.

This sense of timeliness is both fuzzy and arbitrary, absent any firm clinical or methodological guidelines, but the systematic association between timing and effectiveness means that some attempt to control for this dynamic is essential. Better data can improve upon this approach in future studies.

Appendix 4 Pre-implementation trends

4.1 Parallel trends assumption

We checked the underlying assumption that trends in outcome were comparable at different sites in the years prior to PMS implementation. We present those data for the pre-implementation years, by site, for LOS in Figure S4. Our interpretation of these data is that trends are broadly comparable. In terms of level, all sites are very similar. In terms of trend, site 3 is a concern due to the growing variation over time, although the overall trend remains quite similar.

4.2 Anticipatory effects

We checked the underlying assumption that trends in outcome were not changing in the years prior to PMS implementation, perhaps in anticipation of the implementation. We present those data for the years prior to implementation (=0), by site, for LOS in Figure S5. Our interpretation of these data is that trends are broadly comparable. Again, site 3 is the most obvious concern insofar as the trend is changing slightly in the year prior to implementation. However all four outcome lines were increasing modestly in the year(s) prior to implementation.

4.3 Conclusion

We retain all four sites in primary analysis. We perform sensitivity analysis without site 3, due to potential issues with secular trends.

© Annals of Palliative Medicine. All rights reserved.
https://dx.doi.org/10.21037/apm-23-479

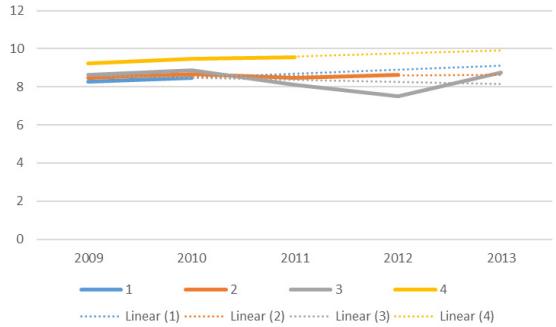


Figure S4 Pre-PMS implementation trends for LOS, by site.

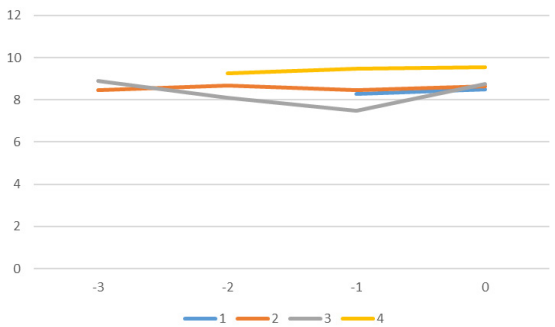


Figure S5 Pre-PMS implementation trends for LOS, by site.

References

1. Health Service Executive. Palliative Care Services: Three-year development framework. Dublin: HSE Primary Care Division; 2017.
2. Irish Hospice Foundation. A Baseline Study on the Provision of Hospice/Specialist Palliative Care Services in Ireland. Dublin 2006. ISBN 0-9545880-3-9.
3. Etkind SN, Bone AE, Gomes B, *et al.* How many people will need palliative care in 2040? Past trends, future projections and implications for services. BMC Med 2017;15:102.
4. Hua M, Li G, Clancy C, *et al.* Validation of the V66.7 Code for Palliative Care Consultation in a Single Academic Medical Center. J Palliat Med 2017;20:372-7.
5. Jagielka P. What Are We Estimating When We Estimate Difference-in-Differences? Impact Evaluations. Washington, DC: World Bank; 2019.
6. Roth J, Sant'Anna PHC, Bilinski A, Poe J. What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. arXiv. 2022.
7. Duhat A, Kondylis F, Loeser J, Piza C. DiD you see Beta? Beta who? Part 1. Impact Evaluations. Vol 2023. Washington, DC: World Bank; 2021.
8. Pinzón E. Econometrics strikes back: GMM and two-way fixed effects. Stata Users' Group Meeting; 2022.
9. May P, Garrido MM, Cassel JB, *et al.* Using Length of Stay to Control for Unobserved Heterogeneity When Estimating Treatment Effect on Hospital Costs with Observational Data: Issues of Reliability, Robustness, and Usefulness. Health Serv Res 2016;51:2020-43.
10. May P, Normand C. Analyzing the Impact of Palliative Care Interventions on Cost of Hospitalization: Practical Guidance for Choice of Dependent Variable. J Pain Symptom Manage 2016;52:100-6.

© Annals of Palliative Medicine. All rights reserved.
https://dx.doi.org/10.21037/apm-23-479