# National, centralised hospital datasets can inform clinical trial outcomes in prostate cancer: a pilot study in the STAMPEDE trial

Harriet P. Mintz[1,2], Felicity Evison[2], Helen M. Parsons[1], Matthew R. Sydes[3], Melissa R. Spears[3], Prashant Patel[2,4], the STAMPEDE trial group, Nicholas D. James[2,4]

[1] Warwick Medical School, University of Warwick, Coventry, England; [2] University Hospital Birmingham NHS Foundation Trust, Birmingham, England; [3] MRC Clinical Trials Unit at UCL, London, England; [4] Institute of Cancer and Genomic Sciences, University of Birmingham, England

## AIM

Can routinely collected data be used to capture clinical trial events?

To develop a method to identify these major events in prostate cancer

To test if clustering of routinely collected NHS data correlates with major progressive disease events

## ABSTRACT

**Background:**

Hospital Episode Statistics (HES) are routinely collected data describing National Health Service (NHS) hospital visits in England, with procedure & disease codes. This study, embedded in STAMPEDE, aimed to build a model using HES, linked to primary medical records & trial case report forms (CRFs) to identify progressive disease events (PDEs), including skeletal-related events (SREs).

**Methods:**

Analysis of 5 STAMPEDE patients (pts) in 2 stages (data to Jul 16). 1: Detailed manual note review of 3 pts' PDEs were compared to HES & CRFs to build model. 2: Used model to use HES to identify possible PDEs in 2 pts, verified by note review & compared to CRFs. Created algorithm rules to identify PDEs per 8 week interval plus further analysis of HES coding to find SREs.

**Results:**

Prostate cancer PDEs coincided with clustering of HES events. HES found 4 PDEs omitted from CRFs but missed 2 (total PDEs: HES 10, CRFs 8). HES found a false positive CRF PDE. Compared with note review HES missed 4 PDEs (false negatives), with 2 missed & 2 upgraded to PDEs post-standard query procedures, plus HES found 3 false positives (1 STAMPEDE treatment & 2 delayed treatments post-PDE). Hence HES found 71% of PDEs in note review (HES 10, note review 14). CRFs found 57% of PDEs compared to note review (CRFs 8, note review 14). Hence HES found 14% more PDEs than were recorded in CRFs compared to note review. HES identified 4 additional SREs not recorded in CRFs but missed 4.

**Conclusions:**

Hospital record review revealed site staff may miss reporting major clinical efficacy outcome events on CRFs, especially nearer end-of-life. HES successfully identified most PDEs (often found as a cluster of SREs), plus additional trial events not reported on CRFs compared to note review and as predicted HES & CRFs found less PDEs. PDEs & SREs missed from CRF recording can be identified in HES. This confirmed use of HES to detect PDEs is feasible. HES-identified events have potential as a primary data source when subsequently verified by standard data queries. Future work will test this model prospectively in the forthcoming BladderPath trial. It may offer a superior, cost-effective method of primary data collection compared to traditional CRF recording.

## METHOD – Data collection - Stage 1

### Data analysed

**Routine hospital statistic data**
England NHS Hospital Episode Statistics: Inpatient, outpatient and A&E data incl. disease & procedure codes (1)

**Hospital administration system** Incl. clinical records, correspondence, lab results (clinical note review)

**STAMPEDE Clinical trial Case Report Forms**
Collect information for trial analysis

### Patients analysed

- 5 STAMPEDE patients recruited at Queen Elizabeth (QE) Hospital Birmingham
- Randomised to original arms A-F

## METHOD - Analysis - stage 1

1. **CLINICAL NOTES BASED ANALYSIS - 3 patients**
   In depth case studies using clinical records to identify progressive disease events
2. **CLUSTERING**
   Identified events often found as a cluster of SREs
   Therefore analysis of event clustering was undertaken
3. **MODEL RULES**
   Optimum clustering for max. sensitivity used (8 week intervals)
   Identifying events from output and then assigning model rules
4. **ROUTINE HOSPITAL STATISTICS DATA ANALYSIS**
   Analysed coding to determine which codes identified progressive disease events and SREs
5. **CLINICAL TRIAL CASE REPORT FORMS**
   Compared events from routine data to STAMPEDE CRFs
   Assessed the suitability of routine data for patient follow-up
6. **ROUTINE HOSPITAL STATISTICS DATA ANALYSIS - 2 patients**
   Model rules applied to routine data
   Hypothesised where progressive disease events experienced
7. **MODEL VERIFICATION**
   Compared routine data results to clinical notes and trial CRFs
   Confirmed the model rules
8. **MODEL REFINEMENT**
   Model refinement to identify missed events from SRE coding
9. **CONFIRMED ROUTINE HOSPITAL STATISTIC DATA ACCURACY**
   Used all 3 data sets to build aggregate record of 'true' patient history to compare accuracy

### The STAMPEDE trial

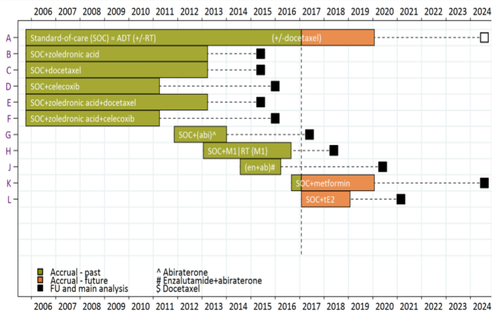
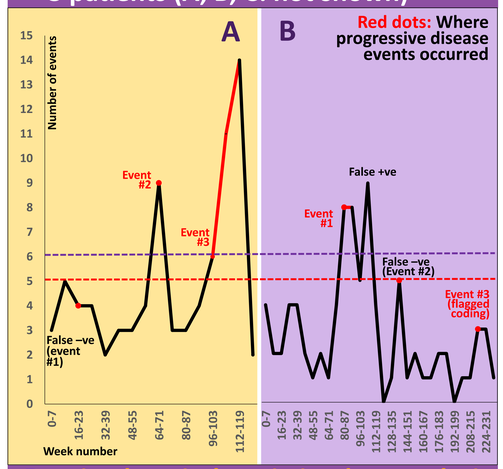



Fig. 1: The STAMPEDE trial; the largest interventional prostate cancer trial in the world. Multi-arm multi-stage currently with arms A-L

## RESULTS - Stage 1: Identifying progressive events

### Medical record review analysis - 3 patients (A, B, C: not shown)




### Routine hospital statistics data analysis - 2 patients (D, E)

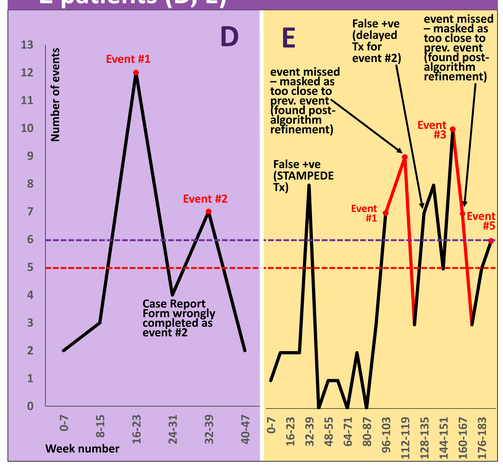



Fig 2: Event clustering from routine hospital statistic data to identify progressive events (red dots); 8 weeks groupings were specified. (A-B: Method 1-5; D-E: Method 6-7)

2 model rules: red line = query progressive event defined by 1st interval > 5 events
purple line = query subsequent progression event only if the number of events from the previous progression drops to < 6 events

Model refining: Routine hospital statistic coding identified SRE in peak <5 events = event

## OVERALL RESULTS - Stage 1: Identifying progressive events

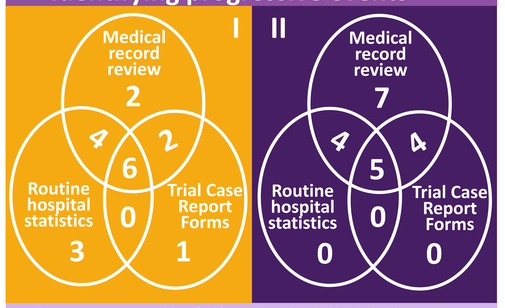



Fig 3: Prostate cancer events identified from routine hospital statistics data
(I) Progressive disease events (PDEs) isolated from clustering;
(II) Skeletal Related Events (SREs) isolated from routine hospital statistic data coding

**Events coincided with clustering of routine hospital statistic data**

**Routine hospital statistics data identified:**

- 71% of PDEs (10/14)
- Including 4 PDEs not in trial CRFs
- An equal number of SREs to trial data (9/20) - 45%
- Including 4 SREs not found on trial CRFs

**Routine hospital statistics data accuracy increased from identifying 31% (4/13) (2007-09) to 71% (5/7) of SREs (2010-14)**

## RESULTS - Stage 1: Routine hospital statistics data accuracy (patients A-C)

| ID | Data | Number events | % accuracy |
|---|---|---|---|
| A | Total confirmed events | 97 | REFERENCE |
| | Routine hospital statistics | 68 | **70** |
| | Medical record review | 91 | 94 |
| B | Total confirmed events | 82 | REFERENCE |
| | Routine hospital statistics | 65 | **79** |
| | Medical record review | 80 | 98 |
| C | Total confirmed events | 20 | REFERENCE |
| | Routine hospital statistics | 16 | **80** |
| | Medical record review | 20 | 100 |

Table 1: Accuracy by aggregating medical record data & routine hospital statistics data. Event = any confirmed prostate cancer event at QE Hospital e.g. surgery, chemo, radiotherapy, imaging (excl. X-rays-not coded for & non-specific to oncology)

**Routine data: 75% accuracy**

## RESULTS - Stage 2 (in progress)

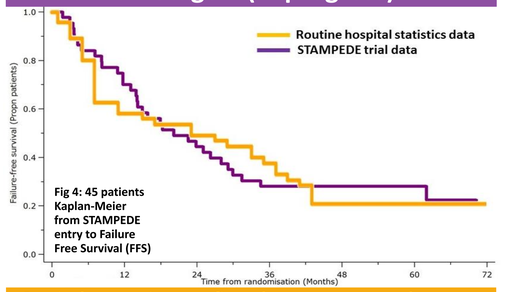



Fig 4: 45 patients Kaplan-Meier from STAMPEDE entry to Failure Free Survival (FFS)

**Survival comparison confirmed event agreement between trial data and the v1.0 algorithm utilising routine data**

## FUTURE WORK

**Stage 2 (In progress)**

- Create an algorithm using the 5 Stage 1 patients in R (3) to automatically identify events
- Check this reproduces the manual Stage 1 output
- Validate algorithm on a further 45 STAMPEDE Queen Elizabeth patients
- Further refine algorithm

**Stage 3**

- Further validate algorithm in ~150 STAMPEDE Queen Elizabeth hospital patients

**Stage 4**

- Validation in all English STAMPEDE patients on Arms A-F

## IMPLICATIONS

**Reduce loss to follow-up**

Allow follow-up of catastrophic events that may go unreported when a patient has not returned to clinic. Ease burden on hospitals reporting trial outcomes; with increased accuracy

**Assess new drugs: validate trial results & health economics**

More events recorded, hence more analysis undertaken incl. costs associated, enabling routine access to new drugs

**Pseudo-trials**

Faster, cheaper clinical trials on population data

## CONCLUSION

Clustering identified progression events

Routine hospital statistic data:

- detected missed events
- identified more events than traditional trial data
- Identified the majority of events
- aided the collection of SRE data
- missed some events however identified more than trial data

For this purpose routine data seems to have sufficient accuracy

Algorithms seem feasible to detect progression events and SREs in prostate cancer

## CONTACT

Hattie Mintz

@HattieMintz
warwick.ac.uk/hattiemintz
H.P.Mintz@warwick.ac.uk

## REFERENCES

1) HSCIC. Hospital Episode Statistics- http://content.digital.nhs.uk/hes (accessed 28 October 2016)

2) Grieve R, Abrams K, Claxton K, Goldacre B, James N, Nicholl J, Parmar M, Parker C, Sekhon JS, Smeeth L, Spiegelhalter D. Cancer Drugs Fund requires further reform. BMJ 2016;354:i5090

3) R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.