Report

Signal detection of spontaneous medical device reports over time accounting for multiple comparisons

Lan Kelly, Ty Stanford, et al.

Table of contents

1 Abstract **2**

2 Introduction **3**

3 Methods **4**

3.1 Data aquisition 4
3.2 Analysis data 9
3.3 Signal detection over time 9
3.3.1 Proportional reporting ratio (PRR) 10
3.3.2 Bayesian Confidence Propagation Neural Network (BCPNN) Information Component (IC) 11
3.3.3 maxSPRT 12
3.4 Analysis choices 13

4 Results **14**

5 Conclusion **18**

6 Session information **19**

1 Abstract

2 Introduction

Adverse events from implantable medical devices are commonly reported to regulatory bodies in the form of unstructured free text in spontaneous reports. Detecting safety signals from the reports for post-market surveillance can be challenging. Spontaneous reports of adverse events may be submitted by manufacturers, clinicians and consumers to the Database of Adverse Event Notifications (DAEN), maintained by the Australian Therapeutic Goods Administration (TGA) in the form of unstructured free text. Pelvic (urogynaecological) mesh was used to treat women for pelvic organ prolapse and stress urinary incontinence. While the device provided benefits to some women, others experienced serious complications.

The aim of this study was to classify the DAEN reports into related adverse events and detect potential safety signals from use of pelvic mesh.

3 Methods

The DAEN was searched for reports on pelvic and hernia mesh from 2012-2017. Topic modelling, a Natural Language Processing technique, was used to profile the unstructured text into mixtures of clinically relevant topics. A report was considered to contain a particular topic X if the probability it contains words relating to the topic, $\mathrm{P}(\text{topic} = X|\text{ document})$, was over a certain threshold, which was varied in the analysis.

Disproportionality analysis was used to detect potential signals from the most frequent clinical topic from pelvic mesh, with hernia mesh and other devices used as comparators. Measures are based on a 2×2 contingency table for the number of adverse events with and without the most frequent topic in the device of interest and the comparator. Testing was performed on the DAEN at quarterly intervals, if new data were accumulated in the interval, over the study period, commencing in 2012.

3.1 Data aquisition

The data is thanks to curtis-murray at his MedicalDevicesNLP repo

- Natural language processing of the TGA spontaneous reports of medical device database (DAEN)
- Each record has an estimate of P(`topic == "pain"` | `Level`, `Doc`) using hierarchical stochastic block modelling (hSBM)
- P(`topic == "pain"` | `Level`, `Doc`) estimates for each record are roughly interpreted as the proportion of the NLP analysed free text that is considered as using/describing words related to pain

And example record and processing values:

Report_ID	Report date	Class	Device	P(‘pain’|doc)	ARTG no.	Event	Source	Event type	Description
37537	2015-02-06	other_devices	Class IIa	0.029	137859	Injury	Industry	Mechanical	Patient admitted for routine SFA angioplasty. The physician had completed the procedure without incident and had withdrawn the balloon catheter fro…
36797	2015-08-27	other_mesh	Class III	0.020	219240	Injury	Industry	Material	3 weeks post-op, patient contacted the surgeon saying that the wound was opening, and she could see the implant. The patient sent photos, and there…

Report_ID	Report date	Class	Device	P('pain'|doc)	ARTG no.	Event	Source	Event type	Description
40917	2016-04-24	pelvic_mesh	Class IIb	0.538	92718	Injury	Consumer	Other	Have had pelvic pain, pain with sex incontinence with bowel and bladder, cannot sit, walk, stand on feet for extended time.
45432	2017-03-29	hernia_mesh	Class IIb	0.214	98833	Injury	Consumer	Other	Atrium mesh implanted in my abdomen. Severe right sided abdominal pain ongoing. Also had further surgery to repair hernia due to reoccurrence of he…

Report_ID	Report date	Class	Device	P('pain'|doc)	ARTG no.	Event	Source	Event type	Description
44402	2017-12-01	other_device	Class III	0.000	149128	No Injury	Other	Other	Sponsor distributed a Customer Letter dated 20th December 2016 announcing the immediate discontinuation of their CE-marked Umbilical Vessel Cathete...
45624	2017-12-04	other_device	Class 1	0.000	121950	No Injury	Health Professional	Mechanical	Operating table started tilting patient on its own during procedure. Emergency stop button pressed - table stopped briefly and began to tilt patien...

Report_ID	Report date	Class	Device	P('pain'|doc)	ARTG no.	Event	Source	Event type	Description
45265	2017-12-18	other_device	Class IIb	0.250	177101	Injury	Industry	Other	Pain, possible rupture and swelling around prosthesis.

3.2 Analysis data

Signal detection of disproportionate adverse events (AEs) will often have tabulated count data accumulated over time. The data at time point t can be summarised as below:

	AE(s) $\in Y$	AE(s) $\in \bar{Y}$
Target exposure	a_t	b_t
Comparator exposure	c_t	d_t

where

- AE(s) Y is the set of AEs (or singular AE) of interest,
- AE(s) $\bar{Y}$ is the complementary set to the AEs of interest,
- *Target exposure* is the medical device(s) of interest,
- *Comparator exposure* is the medical devices to which the *Target exposure* is being compared, and
- a_t, b_t, c_t and d_t (all $\in \mathbb{Z}^+$) are the respective counts of AEs recorded up until (i.e., cumulative) time t.

In the motivating example of the pelvic mesh device, the contingency table can be written more specifically as

	Pain AEs	Not pain AEs
Pelvic mesh	a_t	b_t
Comparator exposure	c_t	d_t

where

- *AEs pain* is the count of AEs that contain "pain" themes greater or equal to some pre-specified threshold $p_t \in (0, 1)$ as estimated by the hSBM (that is, $P($`topic == "pain"`$|$`Level, Doc`$) \geq p_t)$, and
- *Comparator exposure* can be any relevant set of medical devices to compare the pelvic mesh to (e.g., hernia mesh or all other mesh devices or all other devices).

3.3 Signal detection over time

We will consider the three signal detection statistics below:

- Proportional reporting ratio (PRR),

- Bayesian Confidence Propagation Neural Network Information Component (BCPNN IC with MCMC CIs), and
- the maxSPRT statistic

As signal detection is being undertaken repeatedly as data are being accumulated, alpha spending needs to be considered. The below table classifies the aforementioned signal detection methods by their null hypothesis as well as whether they control for the family-wise error rate (FWER)

Null hypothesis	non-FWER version	FWER version
Ratio of pain AEs to all AEs in target and comparator groups has a ratio of 1	PRR	binary, group sequential maxSPRT
Independence of pain AEs and target group (based on marginal counts)	IC	IC with α-spending scheme

We will demonstrate how the group sequential binary maxSPRT, as described in previous work, is equivalent to a FWER-controlled PRR method of signal detection.

Methods used included the Bayesian Confidence Propagation Neural Network (BCPNN) and the maximised Sequential Probability Ratio Test (maxSPRT) which accounted for multiple testing through alpha spending. The BCPNN was used with and without adjusting for multiple testing. The test statistic for BCPNN is the information criterion (IC) which represents the log2 of the ratio of observed to expected adverse events (0 under the null hypothesis of no association between the topic and pelvic mesh).

maxSPRT was developed for near continuous sequential monitoring (called "group sequential" when monitored at discrete time points or after set accumulatios of events), maintaining the correct overall alpha level. The test statistic is based on the maximized (log-)likelihood ratio statistic which uses the observed and expected (under the null hypothesis) reporting ratio assuming binomial adverse event accumulation. The critical value of the likelihood ratio statistic is determined by the 100(1-)% quantile of possible likelihood ratio statistics from binomial adverse event counts under the null hypothesis over the entire group sequential follow-up.

3.3.1 Proportional reporting ratio (PRR)

The PRR estimate is calculated

$$\widehat{\mathrm{PRR}}_t = \frac{\frac{a_t}{a_t+b_t}}{\frac{c_t}{c_t+d_t}}.$$

In the context of signal detection, an elevated proportional reporting ratio is of concern. Therefore the one-sided hypothesis test $H_0 : \mathrm{PRR} \le 1$ (proportional reporting of the target is less than the comparator) is used and is not rejected until

$$\widehat{\mathrm{PRR}}_t \times \exp\left\{ -Z^*_\alpha \sqrt{\frac{1}{a_t} + \frac{1}{a_t+b_t} + \frac{1}{c_t} + \frac{1}{c_t+d_t}} \right\} > 1$$

at the α level where Z^*_α is the $(1-\alpha)^{\mathrm{th}}$ quantile of the standard normal distribution. The above threshold is equivalent to the lower bound of the approximate $100(1-2\alpha)\%$ confidence interval for a standard two-sided hypothesis test.

3.3.2 Bayesian Confidence Propagation Neural Network (BCPNN) Information Component (IC)

The Information Component (IC) statistic is an estimate of the observed-to-expected ratio of the number of target exposure AEs of interest on the $\log_2$-scale under independence between the target exposure and AEs of interest based on information theory (Bate et al., 1998)

$$\mathrm{IC}_{XY} = \log_2 \frac{P_{X,Y}(a_t+b_t, a_t+c_t)}{P_X(a_t+b_t)P_Y(a_t+c_t)}$$

where $P_X(X=x)$ denotes the marginal probability of an observed count x for the target exposure, $P_Y(Y=y)$ denotes the marginal probability of an observed count y for the AE of interest, and $P_{X,Y}(X=x, Y=y)$ denotes the joint probability.

The BCPNN IC of Noren et al. (2006) uses a Bayesian inference based *maximum a posteriori* (m.a.p.) central estimate of the IC,

$$\widehat{\mathrm{IC}}_t = \log_2 \frac{\mathrm{E}\left[\hat{p}_a\right]}{\mathrm{E}\left[\hat{p}_a+\hat{p}_b\right]\mathrm{E}\left[\hat{p}_a+\hat{p}_c\right]}$$

where p_a, p_b and p_c are the (assumed constant over time) underlying probabilities of the multinomial-distributed observed events a_t, b_t and c_t, respectively (p_d corresponding to the count d_t also included). The underlying probabilities are modelled using Dirichlet priors resulting in a Dirichlet posterior distribution. The one-sided null hypothesis of the joint probability target exposure and AEs of interest is equal or less than the marginal products ($H_0 : \mathrm{IC}_t \le 0$) can be rejected when the α quantile of the Markov Chain Monte Carlo (MCMC) empirical distribution is greater than 0. Similarly to the rejection rule for the PRR, this threshold corresponds to the lower bound of the $100(1-2\alpha)\%$ equal-tailed credible region in a two-sided hypothesis test.

3.3.3 maxSPRT

Kulldorff et al. (2011) outlined that the relative risk (RR) at a given point-in-time for accumulated binary data (that is, “success”/“failure” events or AE of interest or not) of a target group relative to a comparator has the maximum likelihood estimate of

$$\widehat{\mathrm{RR}} = z\frac{C_n}{n - C_n}$$

where

- z is the ratio of the total AEs for the comparator to the total AEs for the target,
- C_n is the count of target exposure AEs in X,
- n is the count of all AEs in X (target and comparator exposure), and
- $n - C_n$ is therefore the count of comparator exposure AEs in X.

In the context of our data, the values z, C_n and n are the quantities $\frac{c_t+d_t}{a_t+b_t}$, a_t and $a_t + c_t$, respectively, at time t.

Therefore the RR maximum likelihood estimate at time t can be re-written

$$\begin{aligned}
\widehat{\mathrm{RR}}_t &= \frac{c_t + d_t}{a_t + b_t} \times \frac{a_t}{c_t} \\
&= \frac{\frac{1}{a_t+b_t}}{\frac{1}{c_t+d_t}} \times \frac{a_t}{c_t} \\
&= \frac{\frac{a_t}{a_t+b_t}}{\frac{c_t}{c_t+d_t}}
\end{aligned}$$

which is the PRR estimate at time t as before.

The (maximised) log-likelihood ratio statistic of $\widehat{\mathrm{PRR}}_t$ (equivalently, $\widehat{\mathrm{RR}}_t$) can be determined calculated as

$$\mathrm{LLR}_t = a_t \ln\left(\frac{a_t}{a_t + c_t}\right) + c_t \ln\left(\frac{c_t}{a_t + c_t}\right) - a_t \ln\left(\frac{a_t + b_t}{a_t + b_t + c_t + d_t}\right) - c_t \ln\left(\frac{c_t + d_t}{a_t + b_t + c_t + d_t}\right)$$

3.4 Analysis choices

- Threshold choose
- How many “looks”
- how to choose alpha spending
- sample size limitations for maxsprt - not an issue now can use MCMC method of `EmpiricalCalibration`

4 Results

Pelvic mesh had the highest proportion of reports including the word "pain", followed by other and hernia mesh (Figure 1).

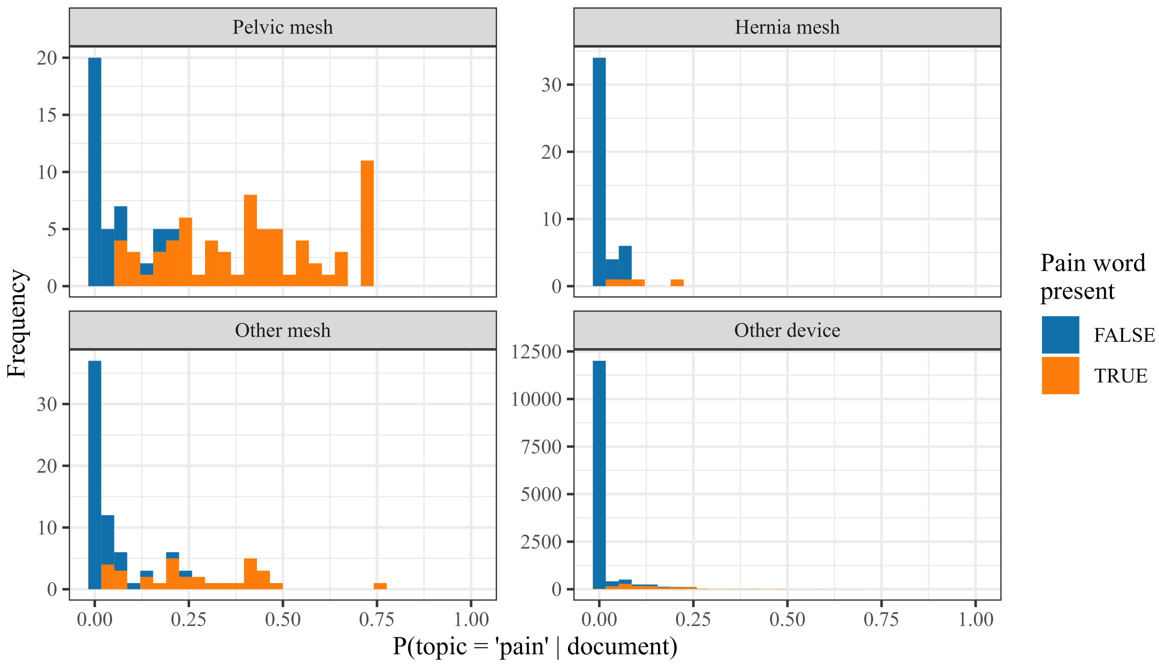

Figure 1: Frequency of pain topic in pelvic, hernia and other mesh and other devices.

The cumulative number of pain topics in pelvic mesh a_t increased sharply in Q4 2014 compared with the previous quarter. The table shows the cumulative 2×2 table each quarter up to $t =$ 2015-Q1 with hernia mesh as the comparator (and pain topic threshold of 0.05).

Table 5: Cumulative quarterly AE counts of pelvic mesh (a_t is pain topic count) compared to hernia mesh (c_t is pain topic count) using a pain topic threshold of 0.05.

Quarter	t	a_t	b_t	c_t	d_t
2013-Q3	1	4	12	1	10
2013-Q4	2	6	14	1	10
2014-Q1	3	6	14	1	11
2014-Q2	4	7	15	1	14
2014-Q3	5	9	17	3	21
2014-Q4	6	26	19	4	27
2015-Q1	7	27	19	4	28
2015-Q2	8	27	19	4	28

Quarter	t	a_t	b_t	c_t	d_t
2015-Q3	9	27	20	4	28
2015-Q4	10	27	20	6	28
2016-Q1	11	30	21	6	28
2016-Q2	12	34	21	6	28
2016-Q3	13	34	21	7	33
2016-Q4	14	36	23	7	33
2017-Q1	15	45	23	8	34
2017-Q2	16	58	24	8	37
2017-Q3	17	68	24	8	38
2017-Q4	18	77	25	8	38

Words associated with "pain" comprised the most frequent clinical topic for pelvic mesh Optimal threshold for P(topic = "pain"| document) pain topics in pelvic mesh for earliest detection from maxSPRT and multiple adjusted BCPNN, depending on comparator (Figure 2):

- Hernia mesh < 0.09
- Other mesh 0.06 – 0.08
- All other devices: Unstable for thresholds > 0.06

Combining results, the best threshold is 0.06 Earliest time period when signal was detected with and without adjustment for multiple testing and maxSPRT by year, with hernia mesh as a comparator:

BCPNN	August 2014
BCPNN with adjustment for multiple testing	December 2014
maxSPRT	December 2014

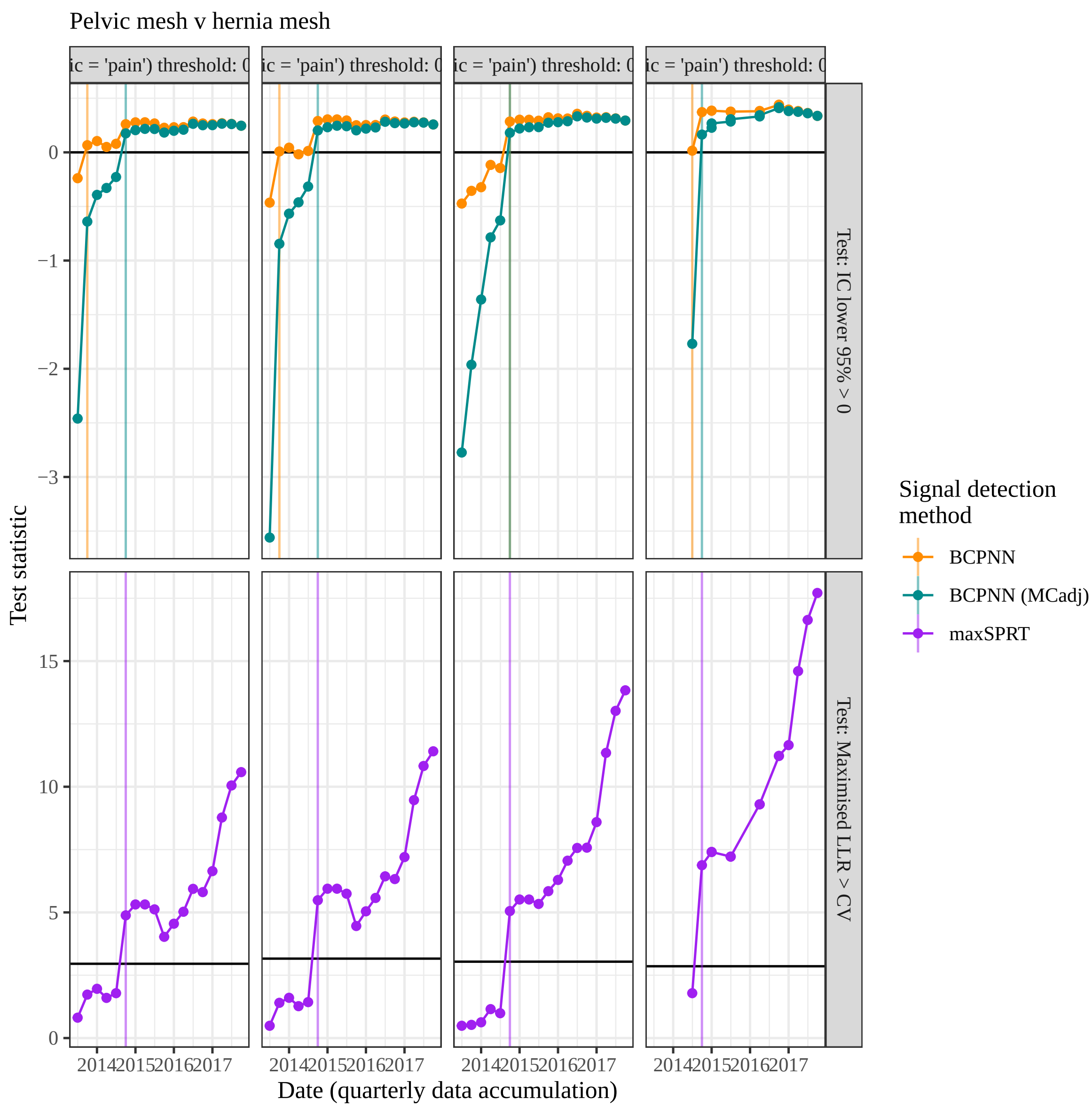

Figure 2: Test statistic values over time.

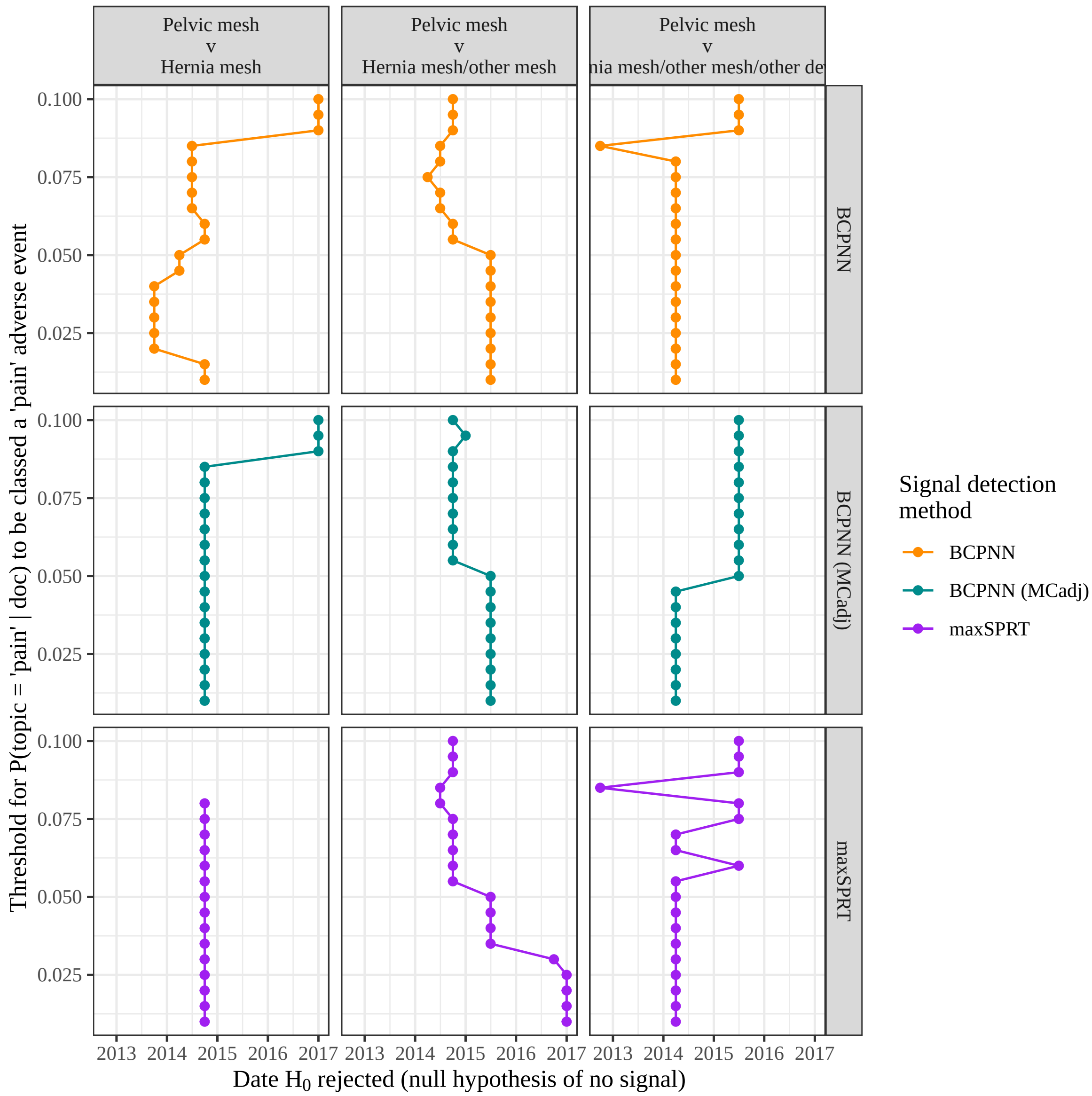

Figure 3: Time when methods reached critical values for signal detection.

5 Conclusion

Urogynaecological mesh was withdrawn from the Australian market in 2018, while our retrospective analysis with a 0.06 threshold detected signals between August – December 2014. We have demonstrated the potential of using topic modelling in spontaneous reports for signal detection in post-market surveillance.

6 Session information

```
format(Sys.time(), '%d %b %Y')
```

```
[1] "25 Sep 2023"
```

```
sessionInfo()
```

```
R version 4.2.2 (2022-10-31 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default

locale:
[1] LC_COLLATE=English_Australia.utf8  LC_CTYPE=English_Australia.utf8
[3] LC_MONETARY=English_Australia.utf8 LC_NUMERIC=C
[5] LC_TIME=English_Australia.utf8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] arrow_11.0.0.2  gsDesign_3.4.0  knitr_1.42      ggrepel_0.9.3
 [5] ggthemes_4.2.4  ggplot2_3.4.1   stringr_1.5.0   lubridate_1.9.2
 [9] forcats_1.0.0   tidyr_1.3.0     dplyr_1.1.2     readr_2.1.4

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.10        pillar_1.9.0       compiler_4.2.2     tools_4.2.2
 [5] bit_4.0.5          digest_0.6.31      jsonlite_1.8.4     evaluate_0.20
 [9] lifecycle_1.0.3    tibble_3.2.1       gtable_0.3.1       timechange_0.2.0
[13] pkgconfig_2.0.3    rlang_1.1.1        cli_3.6.0          rstudioapi_0.14
[17] yaml_2.3.7         xfun_0.37          fastmap_1.1.1      withr_2.5.0
[21] systemfonts_1.0.4 generics_0.1.3     vctrs_0.6.3        hms_1.1.2
[25] bit64_4.0.5        grid_4.2.2         tidyselect_1.2.0   glue_1.6.2
[29] R6_2.5.1           textshaping_0.3.6 fansi_1.0.4        rmarkdown_2.20
[33] farver_2.1.1       tzdb_0.3.0         purrr_1.0.1        magrittr_2.0.3
[37] scales_1.2.1       ellipsis_0.3.2     htmltools_0.5.6    assertthat_0.2.1
[41] xtable_1.8-4       colorspace_2.1-0   ragg_1.2.5         labeling_0.4.2
[45] utf8_1.2.3         stringi_1.7.12     munsell_0.5.0
```