The impact of open access publishing on use and users: Nucleic Acids Research

Professor David Nicholas and Paul Huntington
CIBER
UCL Centre for Publishing
www.publishing.ucl.ac.uk

A presentation about…

- A journal – NAR, which changed
- A methodology – deep log analysis , which evaluates change
- A publishing model – open access, which the information and publishing communities watching with much interest
- Sheer complexities of measuring the impact of OA
- It is just the highlights

Background

- Heard a lot about the impact of open access (OA) publishing on authors, very little about *users*
- Surely OA is about *access* and unless it delivers in this regard then a good part of its attraction goes for the scholarly community
- Therefore we would *expect* greater usage and a *wider* readership

Determining whether this is true

- Have methodology – DLA, to monitor in detail and on a massive scale what goes on in cyberspace. First time employed on single journal; first time on OA.
- Involves refining digital fingerprints & relating to user data. Raises the questions that need to be asked.
- Central tenant of CIBER philosophy is when a digital service is rolled out things never go quite as planned.
- Methodology that produces the evidence-base that surprises, and this study no exception!

Nucleic Acids Research: the journal

- As scholarly as it gets
- Authoratative: impact factor 7.26.
- Ranked by ISI as one of the ten "hottest" journals of the decade.
- Highly specialist with a well defined readership
- Huge journal: 24 issues producing 1500 articles per annum . This is a serious journal.

Nucleic Acids Research: the change

- Moved from a subscription model to a fully open access model in January 2005
- However, not as dramatic a change as it might have been as much of the content was free anyway. Exception last six month's issues
- Has to be taken against a background of growing usage, largely driven by opening the site to search engines in 2003 and use via gateway sites like PubMed and PMC

Main impacts of OA on use/users

1. On article usage
2. On types of articles
3. On types of user

Usage: not so straightforward

Robots: 2% of article views in 2003; by May 2004, 42% of usage; 2005, 20%

Double counting (viewing same article in HTML + PDF)

Very seasonal - usage dipping in summer (and during vacations) and relatively high in the winter and spring academic terms

Leakage to PMC (following slide)

The PMC factor

- Prior to Jan 05 to get NAR full-text PMC users had to go to the OUP site, not necessary after that date Thus from Jan 05 PMC usage of NAR no longer showed up in OUP logs. This meant a leakage of about 2-3% at least. This would undermine ‘the before and after’ analysis so PMC data was stripped out.

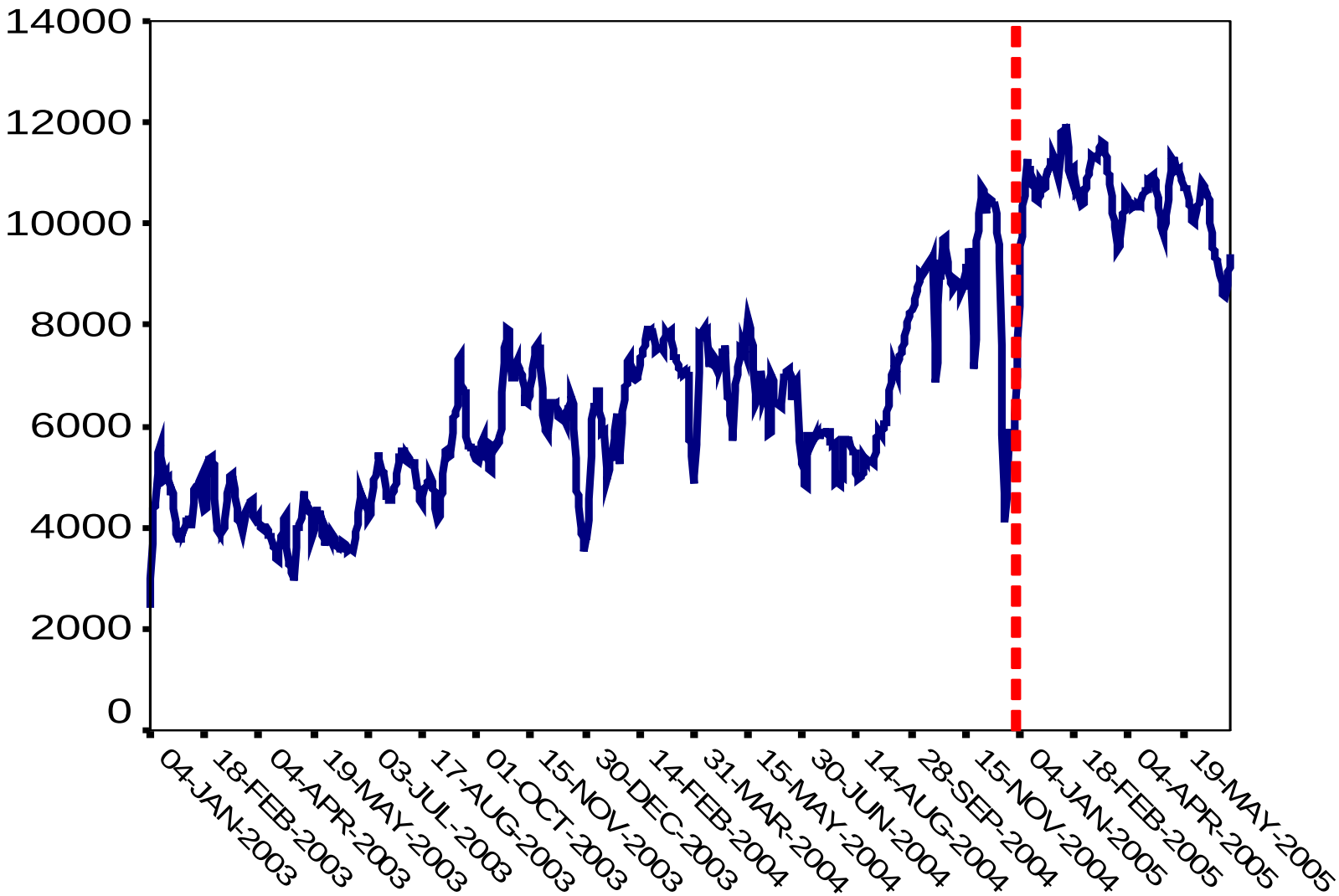
Usage: the big picture

- January 2003 to June 2005.
- 1,500,000 separate internet protocol (IP) numbers used NAR during this period;
- Conducted 5,758,521 million sessions
- Viewed 13,659,000 unique abstract or articles.
- Use was split 57/43 between subscribers and non-subscribers

Monthly usage over the survey period

- Big ‘audited’ growth. 132,000 articles early 2003, rose to 321,000 Jan 2005. 143% increase. Strong especially in the second half of 2004.
- Highest daily article use over the two and half years was recorded for the 9 Feb 2005 (17,151 downloads) - 6,000 more than the daily average for Feb (11,049)
- Clearly not just OA as growth has been going on for some time.

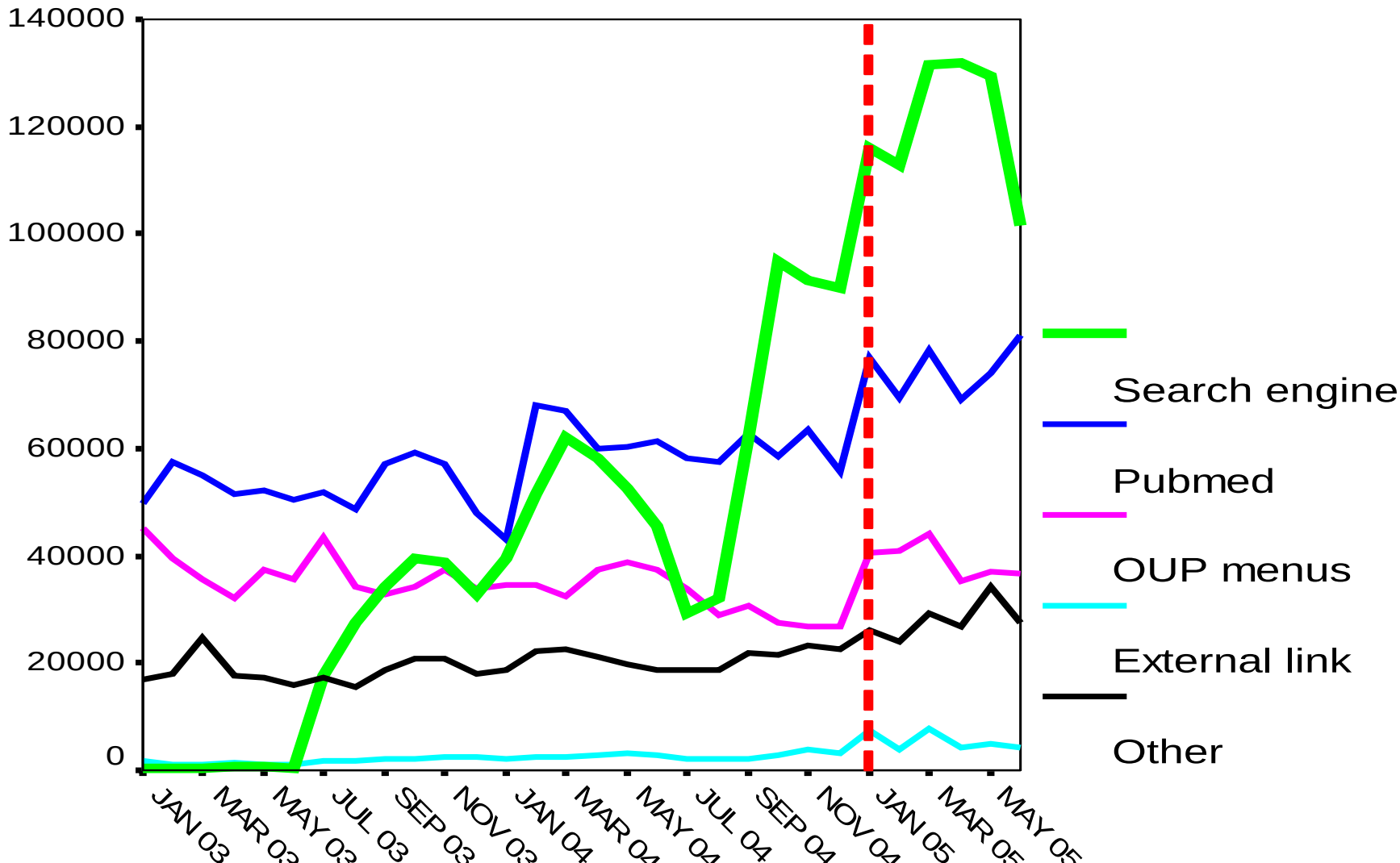
Daily article views for two & half years



So what caused usage to increase by 143% between Jan ’03 to Jan ‘05

- Before June 2003 users could not access or find NAR via a search engine.
- OUP allowed Search Engines and Robots to access the site from June 2003
- This opened the gates to the Google generation

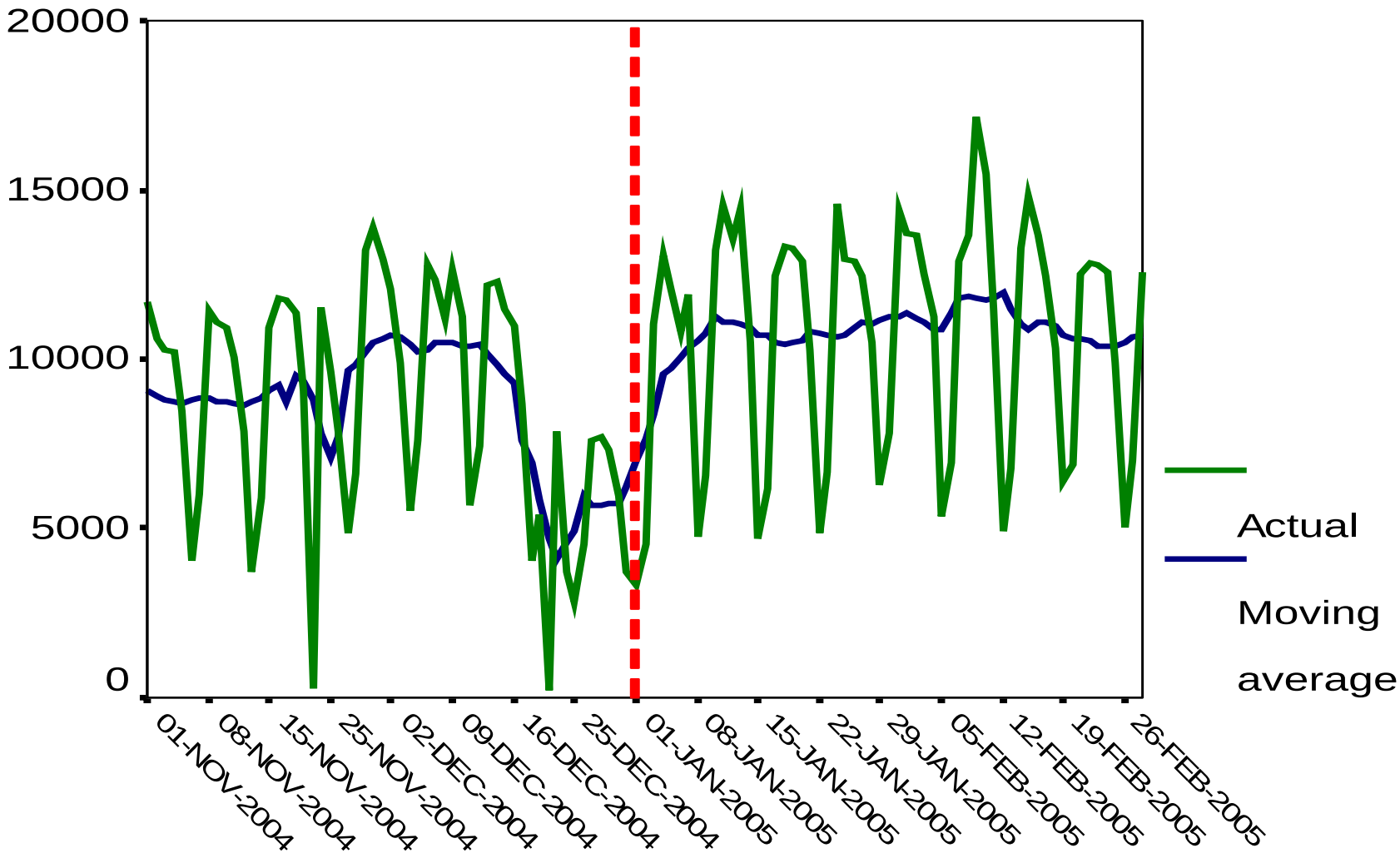
Monthly articles viewed by referrer



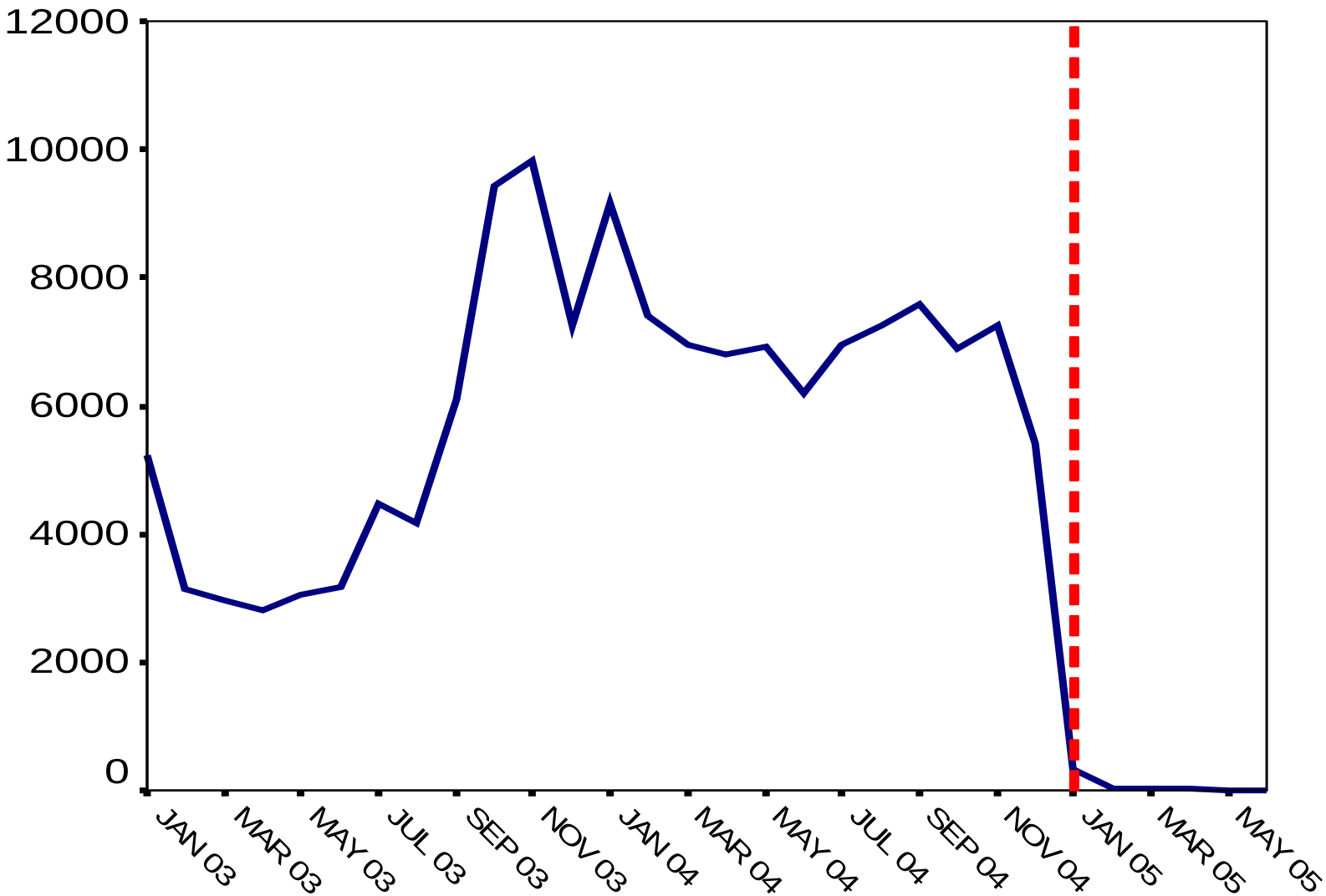
OA impact 1: small impact in use

- Increase in daily average usage of about 19% between Nov 04 to Feb 05 and this probably results from OA. Following slide
- Also Nov to Feb is seasonally a growth period. Thus between Nov ’03 to Feb 04 article use increased by about 11%.
- There is an OA impact but the general increase in usage and the fact that the period is seasonally a growth period obscures result.
- PMC shows considerable growth as a result f-text

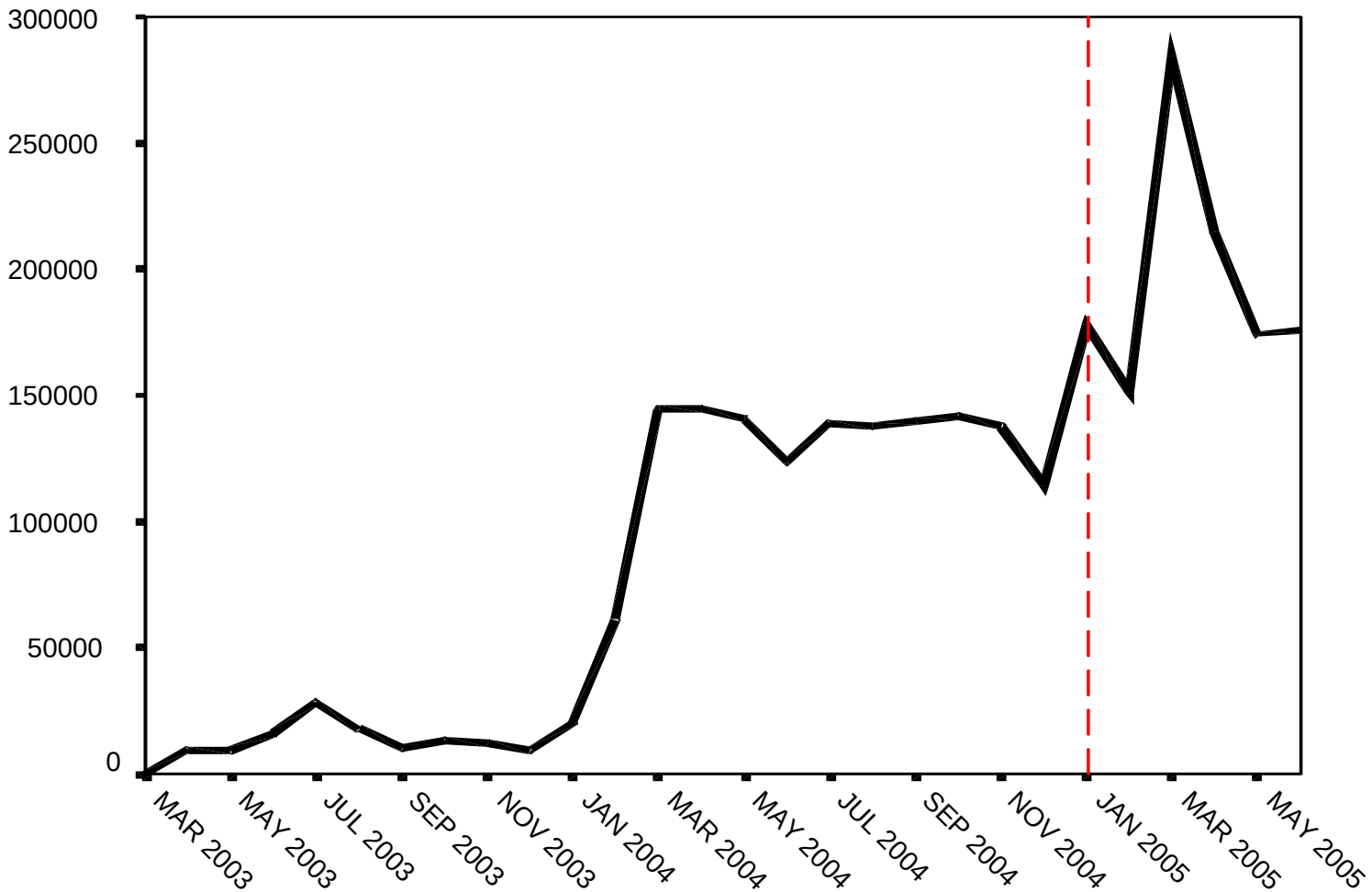
Daily article downloads Nov 04 to Feb 05



PMC article views on NAR site



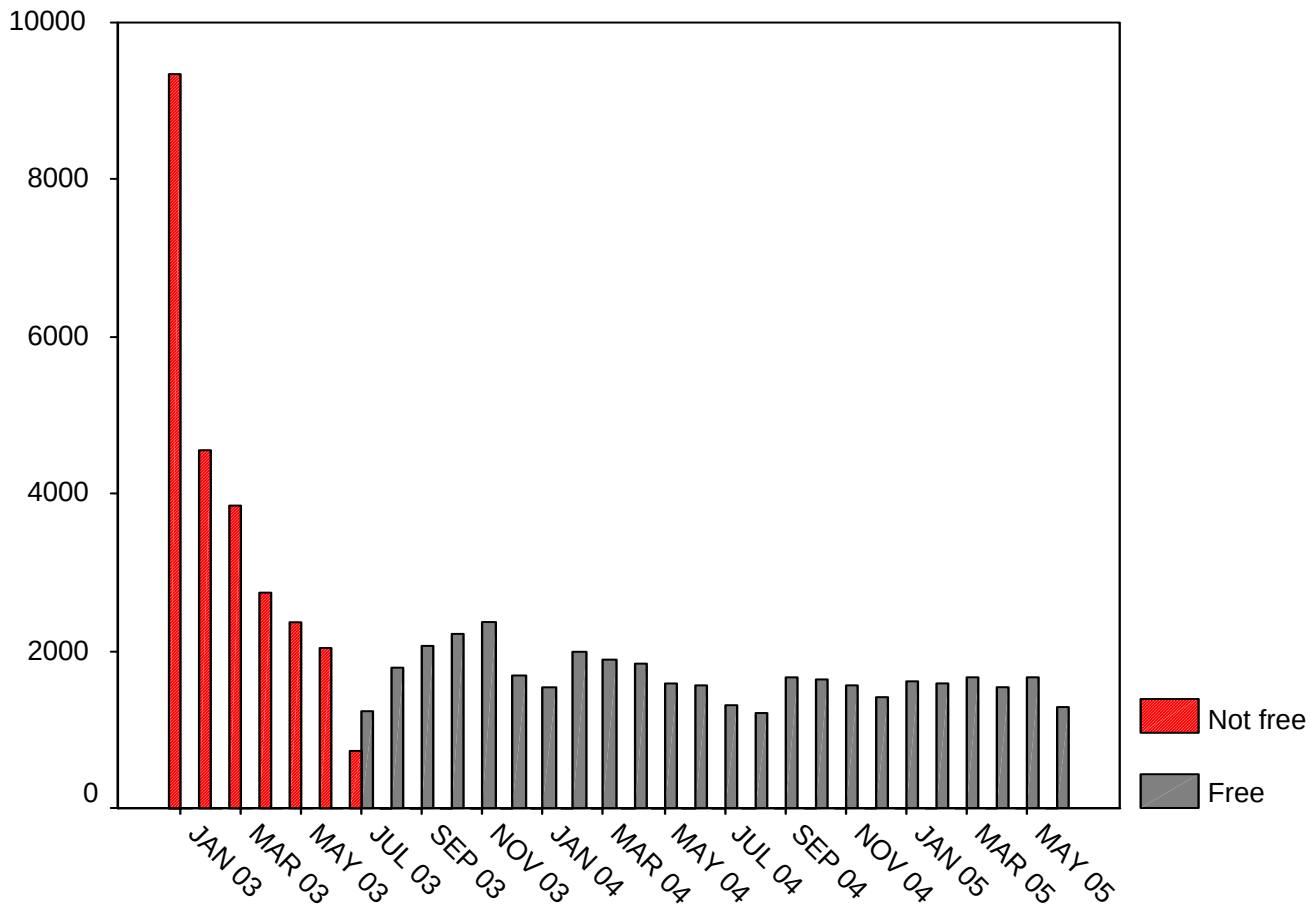
NAR article usage on PubMedCentral – PMC site statistics



OA Impact 2 – on current articles

- The diamonds in the mine.
- Rest of the content free

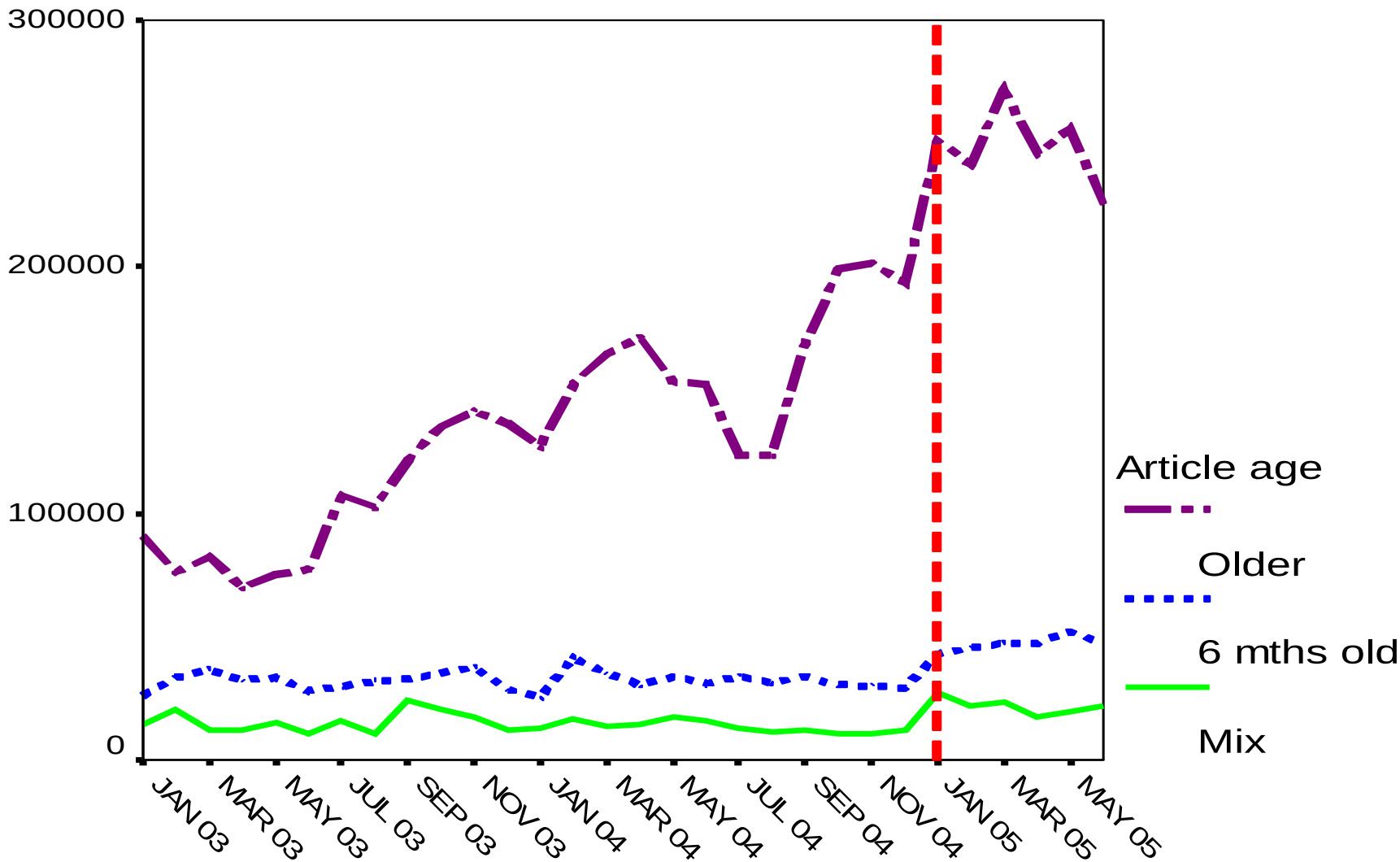
Article views by month (03 to 05) of NAR 31(2), 10th of January 03



Increase in the use of current articles

- During an NAR session users could view articles older than 6 months, articles 6 months and younger or a mixture.
- Before OA articles 6 months and younger were embargoed. *Older articles were free.*
- Between Nov/Dec ’04 to Jan/Feb 05 sessions viewing articles 6 months or younger increased by about 90% (mixed session) and 38% (younger article sessions).

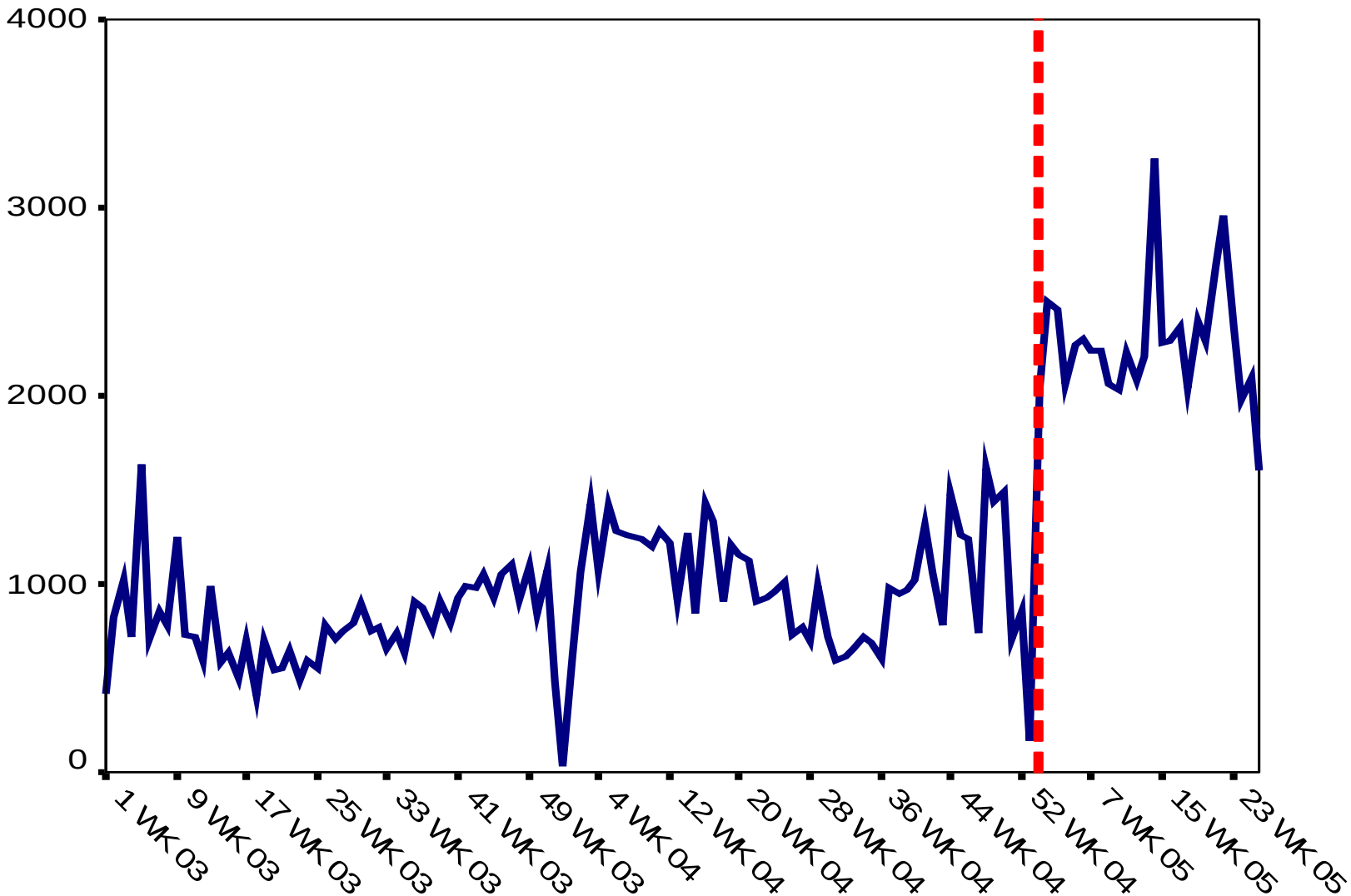
Article views by age



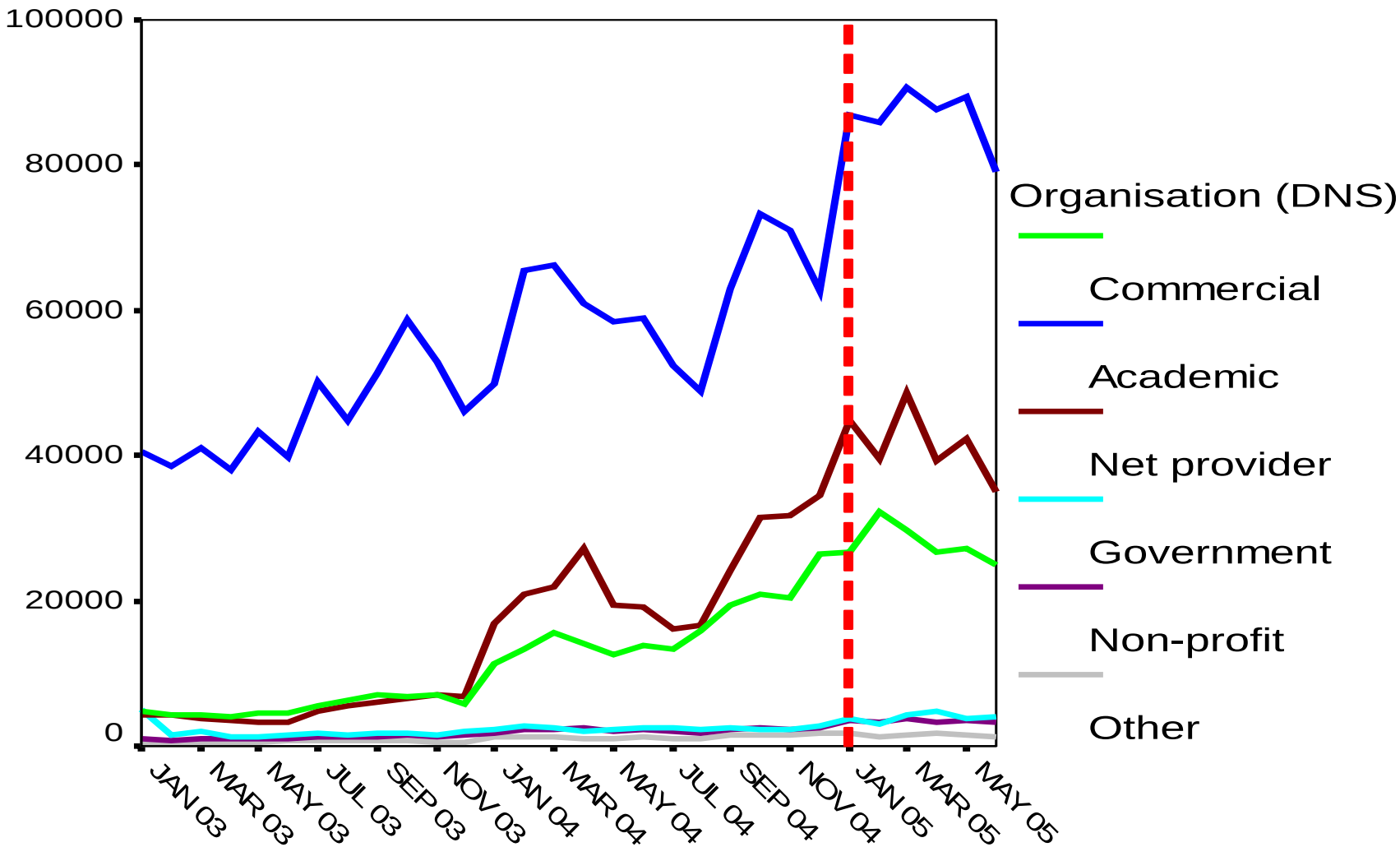
OA Impact 3 – Increased use from East European bloc countries

- Country analysis based on reverse DNS lookup then countries put into regions
- Regions were examined one by one but only Eastern European countries showed an impact
- However seasonal use by these countries is higher in Jan and Feb anyway.
- Nevertheless we estimate an OA effect on usage as an increase of 20%.

Article use – East European bloc countries



Article use by organisation



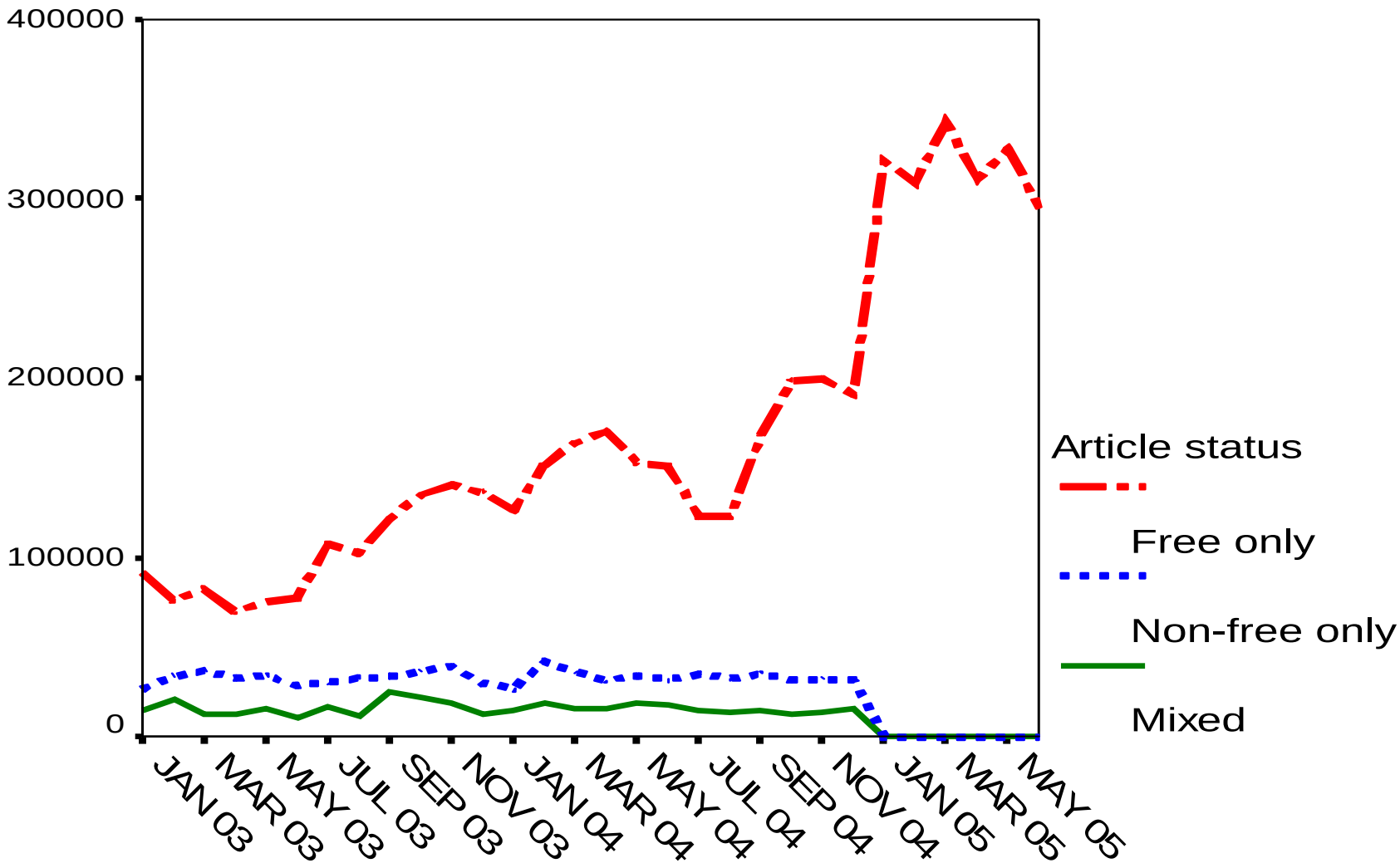
OA in context

1. Free has always been popular
2. Biggest impact result of opening up of the site to search engines.

Free to view articles

- For all three years the top four articles viewed were free articles.
- The most downloaded articles were not the most cited ones. This is partly because of the continuing popularity of *The Molecular Biology Database Collection*, which came first for all three years in usage terms and obtained very few citations.

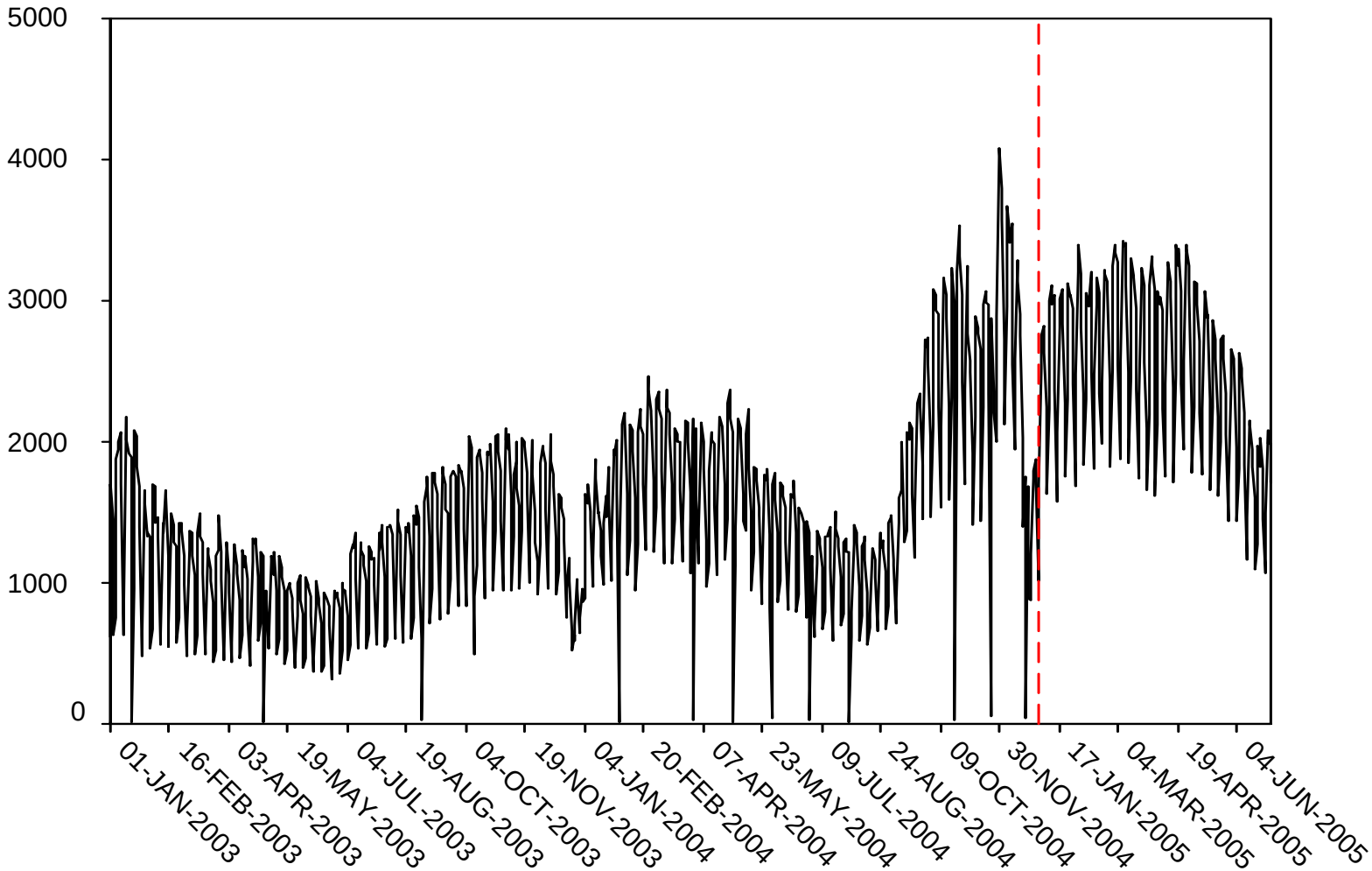
Free to view articles



New user influx –search engine users

- Unprecedented surge in new users in the autumn/early winter of 2004
- This corresponds to the growth period in search engine users
- Search engine users are far more seasonal and drop off rapidly towards the summer – this may suggest undergraduate use.

Daily distribution of new IP numbers accessing the NAR site



Developing hypotheses

- *Hypothesis 1*: non registered users will in the medium term increase their usage by 20% as a result of OA (i.e. East European effect will be repeated).
- *Hypothesis 2*: OA will result in the growth of users just viewing current material to keep up to date.
- *Hypothesis 3*: Those finding the site via a search engine will increase in the second half of ‘05.
- *Hypothesis 4*: Greater use by under graduates will result in a more pronounced seasonal pattern.
- *Hypothesis 5:* NAR is not unique and as we move to other journals we shall find similar