Supplementary material of *Misidentified biomedical resources: Journal guidelines are not a quick fix*

Wytske M. Hepkema, Serge P.J.M. Horbach, Joyce M. Hoek, Willem Halffman

Table of Contents

Supplementary Materials & Methods

1. Journal guidelines

1a Nature Portfolio guidelines regarding cell line authentication

"Cell lines: We strongly encourage deposition of new cell lines in repositories that will distribute them with certificates of authentication. Alternatively, we recommend that authors establish a profile of their new cell lines to allow future authentication. The distribution of human cell lines used in research should not be hindered by restrictions from donors. Researchers developing cell lines must investigate and disclose any restrictions associated with the tissue they are using (see this Nature Editorial for further explanation.) Cell line misidentification and cross-contamination is a common problem with serious consequences. Authors are asked to report on the source and authentication of their cell lines (relevant resources are listed under Further Reading)"

(As downloaded from https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#furthercell on 13-04-2021.)

1b Biomed Central guidelines regarding cell line authentication

"Cell line authentication
If human cell lines are used, authors are strongly encouraged to include the following information in their manuscript:

- The source of the cell line, including when and from where it was obtained
- Whether the cell line has recently been authenticated and by what method
- Whether the cell line has recently been tested for mycoplasma contamination

Further information is available from the International Cell Line Authentication Committee (ICLAC). We recommend that authors check the NCBI database for misidentification and contamination of human cell lines."

(As downloaded from https://www.biomedcentral.com/getpublished/editorial-policies#StandardsofReporting on 13-04-2021.)

1c Nature Portfolio guidelines regarding antibodies

"Nature Portfolio supports the Resource Identification Initiative, with the aim of promoting unique, persistent identification and tracking of key biological resources, including antibodies, cell lines, model organisms and tools. We encourage authors to include unique identifiers provided by the Resource Identification Portal, (RRIDs; for example, Antibody: RRID:AB_2140114; Organism: RRID:MGI_MGI:3840442), in the manuscript. More information on how to include listed RRIDs or generate new RRIDs can be found on the Resource Identification Portal."

(As downloaded from https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#furthercell on 13-04-2021.)

1d BioMed Central guidelines regarding antibodies

"Resource identification
To enable effective tracking of the key resources used to produce the scientific findings reported in the biomedical literature, authors are expected to include a full description of all resources with enough information to allow them to be uniquely identified. In support of the Resource Identification Initiative (RII), we encourage authors to use unique Resource Identifiers (RRIDs) within their manuscript to identify their model organisms, antibodies, or tools."

(As downloaded from https://www.biomedcentral.com/getpublished/editorial-policies#StandardsofReporting on 13-04-2021.)

1e Nature Portfolio guidelines regarding the ARRIVE guidelines

"For primary research manuscripts in the Nature Portfolio journals (Articles, Letters, Brief Communications, Technical Reports) reporting experiments on live vertebrates and/or higher invertebrates, the corresponding author must confirm that all experiments were performed in accordance with relevant guidelines and regulations. The manuscript must include a statement identifying the institutional and/or licensing committee approving the experiments, including any relevant details. Sex and other characteristics of animals that may influence results must be described. Details of housing and husbandry must be included where they are likely to influence experimental results. We recommend following the ARRIVE reporting guidelines when documenting animal studies (PLoS Bio 8(6), e1000412,2010). We also recommend consulting the American Veterinary Medical Association (AVMA) Guidelines for the Euthanasia of Animals (2020), as a comprehensive resource for guidance on veterinary best practice for the anesthesia and euthanasia of animal."

(As downloaded from https://www.nature.com/nature-portfolio/editorial-policies/ethics-and-biosecurity on 13-04-2021.)

1f BioMed Central guidelines regarding the ARRIVE guidelines

"BMC advocates complete and transparent reporting of biomedical and biological research. Please refer to the Minimum standards of reporting checklist when reporting your research (published in BMC Biology). Exact requirements may vary depending on the journal; please refer to the journal's submission guidelines. We also strongly recommend that authors refer to the minimum reporting guidelines for health research hosted by the EQUATOR Network when preparing their manuscript, and FAIRsharing.org for reporting checklists for biological and biomedical research, where applicable. Authors should adhere to these guidelines when drafting their manuscript, and peer reviewers will be asked to refer to these checklists when evaluating such studies.
Checklists are available for a number of study designs, including:

- Randomized controlled trials (CONSORT) and protocols (SPIRIT)
- Systematic reviews and meta-analyses* (PRISMA) and protocols (PRISMA-P)
- Observational studies (STROBE)
- Case reports (CARE)
- Qualitative research (COREQ)
- Diagnostic/prognostic studies (STARD and TRIPOD)
- Economic evaluations (CHEERS)
- Pre-clinical animal studies (ARRIVE)"

(As downloaded from https://www.biomedcentral.com/getpublished/editorial-policies#StandardsofReporting on 13-04-2021.)

2. Detailed description of data obtained through the search method described by Horbach and Halffman [1]

2a Journal selection

In order to assess the effect of journal guidelines regarding cell line authentication and to control for the effect of changes over time, such as a growing awareness of misidentification problems in research communities, we selected for journals with guidelines implemented in the same year. Which resulted in the journals of *BioMed Central (BMC)* and *Nature Portfolio*. *BMC* implemented their guidelines in 2015 and *Nature Portfolio* in 2013, which became stringent in 2015. We did not include all *BMC* and *Nature Research* journals in our analysis. Since we used Web of Science (WoS) for our analysis, this excluded journals not in WoS. To exclude journals not working on cell lines and avoid contamination of the data set, we further refined this list by selecting for relevant research areas, as did Horbach and Halffman [1]. Some journals were discontinued before the implementation of the guidelines in 2015, or were established afterwards. Since it is not possible to see an effect of the guidelines in these journals, we only included journals that published articles in all years between 2014 and 2018.

Since the number of articles on misidentified cell lines can differ per research field, we selected comparable journals as a control based on their citing behaviour. We decided that to select journals working within the same research field, citing behaviour would be the best criteria. Therefore, we did not take other factors, such as prestige, into account when selecting comparable journals. To select comparable journals, we created bibliographic coupling networks based on citing publications in the period 1995 – 2018. To account for differences in the number of authors per paper, we used bibliographic coupling based on fractional counting, for details see Perianes-Rodriguez, Waltman and Van Eck [2].

Bibliographic coupling produces comparable journals for every journal in our set, but since some journals occur more than once in this set, and we wanted a control group of the same size, this required an additional procedure. As a first step, we used the bibliographic coupling to select the 30 most comparable journals for each of the journals of *BMC* and *Nature Portfolio* that we included in our analysis. We then had to select a comparable journal without guidelines for each of these journals from the top 30 without creating doubles. For the first journal on this list (alphabetically sorted on journals with guidelines), we selected the first comparable journal without guidelines regarding cell line authentication. In order to do so, we checked the author guidelines manually and excluded journals with guidelines regarding cell line authentication. In addition, journals with guidelines for RRIDs mentioning cell lines were also excluded. For the second journal and onwards, we selected the first comparable journal without guidelines not already present in our control group. For three journals, we did not find a suitable control in the top 30. Therefore, we excluded these three journals from our analysis.

2b Comparison

We applied the search string as described by Horbach and Halffman [1] in WoS to each journal of our analysis, and listed the number of articles using misidentified cell lines for the years 1995 – 2018 for these journals. Since the yearly number of articles can vary greatly between journals, and the number of articles on cells can vary between journals, we also noted the number of articles with the topic *cell*. We then calculated the number of articles using misidentified cell lines as a percentage of the articles on cells.

Not all journals published articles in all years between 1995 and 2018. To ensure that the data of our control groups best matches the data of the journals with guidelines, we only included the data of the years both the journal with guidelines and the control journal existed. For example, *BMC Cancer* was established in 2001. The control journal of *BMC Cancer* is *Journal of Clinical Oncology*, which was established before 1995. Since we are compared *BMC Cancer* to *Journal of Clinical Oncology*, we only included the data of *Journal of Clinical Oncology* from 2001 onwards.

2c More detailed analysis of *International Journal of Cancer* and *Blood*

We performed a more detailed study for two journals. We chose the *International Journal of Cancer (IJC)* since this journal previously evaluated its guidelines as effective [3]. For the control, we again looked at the top 30 most comparable journals based on citing behaviour (see supplementary material 2a). Of the most comparable journals without guidelines, we picked the first journal with a comparable number of articles using misidentified cell lines, i.e. *Blood*.

For all articles in the *IJC* and *Blood* that used misidentified cell lines as identified with the search string described by Horbach and Halffman [1], we assessed the use of the misidentified cell lines in these papers. We distinguished two categories of false positives: *papers warning about misidentified cell lines* and *papers in which no misidentified cell line was used.* For the articles that did use a misidentified cell line, we distinguished three categories, *misidentified cell line used but aware of problem (1), misidentified cell line used but of same tissue type (2),* and *misidentified cell line used and unaware of problem (3).* For the last two categories we also distinguished between misidentified

cell lines that were used before or after they were first reported to be misidentified. We used the column "misidentification first reported by" of ICLAC's register for this purpose. When there were multiple years in this column we used the first one, when the article was published in the same year as the first reported we counted it as "before", and if there was no year in this column we also counted it as "before" (for exact numbers see supplementary table 4).

The search string described by Horbach and Halffman [1], searches both for the names of misidentified cell lines and for articles that cite the establishing paper of a misidentified cell line. For the articles identified based on the name of the misidentified cell line, we first read the abstract, and in case of doubt the rest of the paper. For the articles found based on the establishing paper, we first checked which establishing paper was cited. Then we checked if the cell line established in that paper was used by reading the material and methods section, in case of doubt the full body of text was searched for this cell line. In addition, if multiple cell lines were used, we also checked those.

Some articles did not state the tissue type of the cell line, or did not mention the specifics of the tissue type, these were counted as *misidentified cell line used, and unaware of problem*. There were also some articles which did not use the misidentified cell line of the cited establishing paper, but did use another misidentified cell line (3 articles in *IJC* and 11 in *Blood*). Since these articles did use a misidentified cell line, they were also counted as *misidentified cell line used, and unaware of problem* or *misidentified cell line used, but of same tissue type*, depending on the cell line. For *Blood*, 48 of the articles were meeting abstracts of which the full texts were unavailable. For one of these, it was clear that the misidentified cell line was mentioned as a warning about misidentified cell lines; the other 47 articles were excluded from our analysis.

3. Detailed description of reanalysis of dataset of Babic, Capes-Davis, Martone, Bairoch, Ozyurt, Gillespie and Bandrowski [4]

Babic, Capes-Davis, Martone, Bairoch, Ozyurt, Gillespie and Bandrowski [4] identified 150.459 articles using cell lines by text-mining the methods section of about two million papers in PubMed central. In short, they used the SciScore tool *Named Entity Recognition-based algorithm* to identify cell-line names in the methods sections. Then, these cell-line names were coupled to ICLAC's list of misidentified cell lines including the partially misidentified cell lines, and all synonyms present in Cellosaurus. They coupled the identified cell-line names to this list in two approaches; *strict* and *loose*. To avoid false positives, we continued only with the strict approach. In addition, we excluded the partially misidentified cell lines, and continued only with the misidentified cell lines of ICLAC's list.

So far as possible, we included the same journals as our analysis of articles found with the search string of Horbach and Halffman [1]. However, since this dataset is limited to articles accessible for text-mining, not all journals we previously included were present in this dataset. Then, to assess the effect of guidelines using SciScore text-mining, we first selected all unique PMIDs of articles using cell lines. We then searched these PMIDs in WoS and noted the number of articles for each year in the period 1995-2018 for the journals of *BMC*, *Nature Portfolio*, and journals without guidelines regarding cell line authentication. Subsequently, we searched the PMIDs of articles that used misidentified cell lines in WoS and again noted the number of articles for each year in the period 1995-2018 for the journals of *BMC*, *Nature Portfolio*, and journals without guidelines. Finally, we calculated the number of articles using misidentified cell lines as a percentage of the articles on cells. In our graph we only included the years in which each group had hundred or more articles using a cell line to prevent a skewed image caused by little data.

4. Detailed description of reanalysis of the datasets of Menke, Roelandse, Ozyurt, Martone and Bandrowski [5]

We downloaded supplementary data 8 (cell lines), supplementary data 7 (antibodies), and supplementary data 1 of Menke, Roelandse, Ozyurt, Martone and Bandrowski [5]. For each of these files

we first copied the column containing the journal names and removed the duplicates. For these three lists of journals, we manually divided all journals in four categories; journals of (1) *BMC*, (2) *Nature Portfolio*, (3) other journals with guidelines (regarding cell line authentication, antibodies, or organisms depending on the file), and (4) journals without guidelines. For cell line authentication, journals with guidelines for RRIDs mentioning cell lines and no separate mention of cell line authentication were counted as journals without guidelines regarding cell line authentication. For antibodies, we did include journals that ask for RRIDs for antibodies. For organism, journals were counted as 'with guidelines' if the ARRIVE-guidelines were mentioned. With these four groups of journals, we labelled each data-entry for these four groups. Then, we sorted the data entries per journal group and year, and calculated the yearly average of percentage of cell lines/ antibodies/ organism that were identifiable/ authenticated. In our graph we only included the years in which each group had ten or more different journals to prevent presenting a skewed image caused by little data.

Supplementary Tables

Supplementary table 1 Cell lines: Identifiability & Authentication as obtained through reanalysis of the dataset of Menke, Roelandse, Ozyurt, Martone and Bandrowski [5]

	BMC					Nature					Other Journals without guidelines					Journals without guidelines				
Year	total number of journals analyzed	total number of articles analyzed	total number of cell lines found	%iden-tifiable	% authen-ticated	total number of journals analyzed	total number of articles analyzed	total number of cell lines found	%iden-tifiable	% authen-ticated	total number of journals analyzed	total number of articles analyzed	total number of cell lines found	%iden-tifiable	%authen-ticated	total number of journals analyzed	total number of articles analyzed	total number of cell lines found	%iden-tifiable	% authen-ticated
1996	0	0	0	0	0	0	0	0	0	0	1	13	22	40.91%	7.69%	0	0	0	0	0
1997	0	0	0	0	0	0	0	0	0	0	3	462	862	37.73%	4.12%	0	0	0	0	0
1998	0	0	0	0	0	0	0	0	0	0	3	484	917	37.05%	4.48%	0	0	0	0	0
1999	0	0	0	0	0	0	0	0	0	0	3	408	768	39.45%	5.57%	0	0	0	0	0
2000	0	0	0	0	0	0	0	0	0	0	3	405	807	33.91%	2.94%	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	2	337	693	32.89%	4.67%	0	0	0	0	0
2002	4	53	85	32%	3%	1	144	287	45.99%	2.67%	3	327	646	41.25%	5.58%	1	22	32	56.25%	0.00%
2003	0	0	0	0	0	1	173	382	36.39%	3.23%	3	362	669	37.70%	2.48%	1	15	24	54.17%	6.67%
2004	9	145	305	37%	3%	1	156	326	41.41%	6.25%	4	368	676	37.75%	2.30%	2	48	63	39.17%	4.55%
2005	10	257	596	43%	3%	1	124	299	37.79%	7.74%	5	377	741	37.30%	6.16%	3	264	520	52.05%	2.95%
2006	16	426	944	46%	5%	1	106	256	45.70%	0.00%	8	433	871	44.46%	7.01%	3	237	454	49.93%	11.24%
2007	15	491	1011	43%	4%	1	126	303	37.29%	0.00%	10	712	1399	39.17%	5.06%	6	333	632	45.03%	4.27%
2008	19	644	1407	43%	5%	1	136	332	36.45%	0.00%	12	1213	2397	41.68%	4.81%	10	473	986	38.75%	2.21%
2009	30	973	2072	44%	5%	8	373	846	36.96%	2.31%	13	1806	3598	39.94%	5.80%	20	740	1429	35.90%	2.24%
2010	34	1374	3062	44%	5%	13	492	1211	39.95%	2.92%	14	2444	5123	41.46%	5.25%	26	1130	2332	38.59%	3.85%
2011	40	1433	3000	44%	5%	18	602	1351	36.91%	3.93%	20	4137	8600	42.80%	5.66%	30	1444	2805	37.72%	5.47%
2012	41	1265	2569	44%	6%	21	890	1973	40.86%	5.02%	27	6367	13363	42.51%	5.26%	52	2181	4270	38.06%	3.61%
2013	46	1467	3126	45%	7%	22	1534	3539	37.44%	6.36%	31	8204	17323	42.59%	5.82%	72	2871	18950	39.38%	5.30%
2014	49	1910	4192	43%	7%	22	1922	4607	39.17%	10.81%	40	8316	17720	41.10%	7.63%	112	4308	8453	35.97%	4.65%
2015	52	1958	4178	42%	5%	25	3472	7957	38.71%	18.11%	42	8914	20516	40.08%	8.90%	119	4551	8785	38.85%	6.35%
2016	47	1645	3419	41%	8%	25	5750	12799	39.69%	26.31%	43	9663	23080	39.80%	8.21%	126	5354	10239	39.63%	6.82%
2017	46	1627	3510	41%	8%	31	7081	16074	37.85%	27.58%	50	10573	24419	41.59%	12.44%	143	7104	13902	38.40%	8.27%
2018	49	2048	4689	41%	10%	30	5826	14052	37.01%	27.77%	62	8528	18602	40.98%	14.18%	158	9396	19039	38.53%	9.42%
2019	45	1701	4244	43%	11%	13	3453	8103	37.87%	28.92%	56	5527	11727	38.92%	12.58%	123	6879	14899	37.81%	11.60%

Supplementary table 2 Cell lines: misidentified as identified by search method of Horbach and Halffman [1]

	BMC			Nature			Journals without guidelines		
Year	Total number of **journals** analyzed with articles with topic cell	Total number of **articles** with topic cell	Total number of articles that use a **misidentified** cell line	Total number of **journals** analyzed with articles with topic cell	Total number of **articles** with topic cell	Total number of articles that use a **misidentified** cell line	Total number of **journals** analyzed with articles with topic cell	Total number of **articles** with topic cell	Total number of articles that use a **misidentified** cell line
1995	7	183	8	16	2309	50	20	5648	61
1996	7	126	1	19	2901	50	23	6759	71
1997	8	160	3	22	3551	56	27	8818	89
1998	8	189	4	24	4340	45	29	9556	107
1999	8	173	2	26	4618	55	31	10513	96
2000	10	251	3	29	4032	68	38	11337	99
2001	15	344	5	30	5165	55	43	13094	108
2002	19	405	2	31	5385	46	47	14416	108
2003	21	361	7	32	5681	48	49	16517	120
2004	23	503	16	33	4845	63	52	18523	125
2005	31	934	13	35	5299	48	62	20193	115
2006	39	1270	14	36	5873	33	69	21706	122
2007	39	1355	13	37	5406	32	71	22065	110
2008	50	1788	19	38	5755	39	83	22706	127
2009	57	2273	24	42	5023	24	94	24987	152
2010	64	3073	29	44	5933	20	103	27796	112
2011	71	3332	33	47	5729	22	110	33679	148
2012	77	3477	19	48	6074	36	118	35416	153
2013	78	3929	22	48	6720	38	119	35282	137
2014	78	4391	35	48	6636	24	119	34812	136
2015	78	4546	33	48	6162	26	119	33280	135
2016	78	4083	25	48	6451	27	119	34770	130
2017	78	4098	19	48	6620	18	119	33035	94
2018	78	4708	34	48	6372	11	119	32948	118

Supplementary table 3 Cell lines: misidentified as obtained through reanalysis of the dataset of Babic, Capes-Davis, Martone, Bairoch, Ozyurt, Gillespie and Bandrowski [3]

	BMC		**Nature Portfolio**		**Journals without guidelines**	
	Total number of articles that use a cell line	Total number of articles that use a misidentified cell line	Total number of articles that use a cell line	Total number of articles that use a misidentified cell line	Total number of articles that use a cell line	Total number of articles that use a misidentified cell line
2000	2	0	0	0	0	0
2001	21	1	0	0	0	0
2002	53	2	139	6	0	0
2003	57	2	168	7	0	0
2004	98	4	156	14	3	0
2005	233	9	116	4	246	7
2006	296	6	94	4	248	5
2007	420	16	126	8	363	10
2008	480	19	136	5	461	9
2009	610	19	275	9	603	15
2010	733	27	280	12	899	14
2011	748	28	289	9	1069	30
2012	676	16	314	14	1365	16
2013	731	17	386	10	1468	34
2014	833	21	400	12	1530	37
2015	801	15	461	12	1685	43
2016	675	10	504	12	1832	46
2017	572	13	453	13	2209	57
2018	30	0	9	0	61	2

Supplementary table 4 Cell lines: misidentified cell lines in *International Journal of cancer* and *Blood*

International Journal of Cancer							
	Unaware - before first reported	Unaware - after first reported	Unaware - unclear before or after first reported*	Same tissue type - before first reported	Same tissue type - after first reported	Same tissue type - unclear before or after first reported*	Aware of misidentification
1995	1	6	1	1	1	0	0
1996	3	5	3	0	1	0	0
1997	7	1	1	1	3	0	0
1998	3	4	1	0	3	0	2
1999	7	3	0	0	1	0	0
2000	3	1	0	1	2	0	0
2001	3	6	2	1	1	0	0
2002	3	2	0	1	1	0	1
2003	3	3	0	1	2	1	1
2004	3	6	0	1	1	0	0
2005	3	4	1	0	3	0	0
2006	2	4	0	0	1	1	2
2007	2	2	1	0	1	0	0
2008	1	1	2	0	0	1	2
2009	1	4	0	0	1	0	1
2010	1	0	0	1	2	0	0
2011	1	1	0	0	0	0	0
2012	0	0	0	0	0	0	2
2013	0	1	0	0	0	0	0
2014	0	0	0	0	0	0	0
2015	1	0	0	1	0	1	0
2016	0	0	0	0	0	0	0
2017	0	0	0	0	0	0	0
2018	0	0	0	0	0	0	0

Blood							
	Unaware - before first reported	Unaware - after first reported	Unaware - unclear before or after first reported*	Same tissue type - before first reported	Same tissue type - after first reported	Same tissue type - unclear before or after first reported*	Aware of misidentification
1995	10	1	0	2	0	0	0
1996	13	0	0	0	0	1	0
1997	12	1	1	1	0	0	0
1998	1	3	0	1	0	0	0
1999	6	6	1	1	0	0	1
2000	3	1	0	0	0	0	0
2001	3	4	0	1	0	0	0
2002	2	3	0	0	0	0	0
2003	3	3	0	3	0	0	0
2004	0	0	1	0	0	0	0
2005	1	3	0	0	0	0	0
2006	0	1	0	0	2	0	0
2007	1	2	0	0	0	0	0
2008	0	1	0	0	0	0	1
2009	0	2	0	0	0	0	0
2010	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	0
2012	0	3	0	0	0	0	0
2013	0	0	0	0	0	0	0
2014	0	0	0	0	0	0	0
2015	0	0	0	0	0	0	0
2016	0	1	0	0	0	0	0
2017	0	0	0	0	0	0	0
2018	0	0	0	0	0	0	0

*In figure 3 the articles for which it was unclear whether they used the cell line before or after it had been first reported were counted as before.

Supplementary table 5 Antibodies: identifiability as obtained through reanalysis of the dataset of Menke, Roelandse, Ozyurt, Martone and Bandrowski [5]

	BMC				Nature				Other Journals with guidelines				Journals without guidelines			
	total number of journals analyzed	total number of articles analyzed	total number of antibodies found	average % identifiable per year	total number of journals analyzed	total number of articles analyzed	total number of antibodies found	average % identifiable per year	total number of journals analyzed	total number of articles analyzed	total number of antibodies found	average % identifiable per year	total number of journals analyzed	total number of articles analyzed	total number of antibodies found	average % identifiable per year
1996	0	0	0	0	0	0	0	0	1	26	26	7.69%	0	0	0	0
1997	0	0	0	0	0	0	0	0	2	699	699	11.75%	0	0	0	0
1998	0	0	0	0	0	0	0	0	2	790	790	11.10%	0	0	0	0
1999	0	0	0	0	0	0	0	0	2	619	619	15.20%	0	0	0	0
2000	0	0	0	0	0	0	0	0	2	649	649	12.58%	0	0	0	0
2001	2	27	27	18.33%	0	0	0	0	2	572	572	12.48%	0	0	0	0
2002	4	56	56	5.74%	0	0	0	0	2	539	539	14.14%	2	229	229	9.56%
2003	5	106	106	16.01%	0	0	0	0	4	613	613	15.43%	2	263	263	8.06%
2004	9	195	195	19.96%	0	0	0	0	3	576	576	18.39%	3	295	295	8.63%
2005	16	515	515	11.90%	0	0	0	0	4	603	603	15.09%	9	533	533	8.35%
2006	27	783	783	12.79%	0	0	0	0	7	678	678	14.94%	12	611	611	13.42%
2007	25	883	889	15.08%	0	0	0	0	11	1156	1157	14.66%	14	839	841	12.69%
2008	32	1215	1218	15.69%	2	31	31	13.25%	14	1922	1934	14.62%	21	1182	1183	11.89%
2009	42	1724	1727	15.14%	12	360	360	15.34%	15	2703	2717	19.03%	32	1770	1773	13.09%
2010	55	2523	2539	16.91%	14	419	419	17.83%	20	3954	3970	18.04%	46	2283	2284	15.95%
2011	59	2520	2534	17.07%	19	620	625	17.59%	30	6551	6581	18.01%	61	2900	2906	14.80%
2012	60	2610	2637	16.98%	21	867	869	20.76%	48	10188	10309	20.87%	106	4540	4550	16.91%
2013	73	3040	3074	19.40%	23	1414	1415	23.25%	63	13209	13289	22.99%	118	6086	6105	17.18%
2014	77	3668	3694	19.59%	23	1934	1944	28.91%	70	12957	13192	23.75%	168	8276	8351	19.29%
2015	82	3755	3808	22.24%	23	4233	4248	30.96%	76	13575	13913	26.33%	179	8785	8830	19.08%
2016	71	3182	3305	22.14%	26	7607	7644	32.40%	81	14640	16178	31.03%	199	10167	10191	19.75%
2017	76	3288	3436	23.07%	29	9270	9470	32.02%	93	17168	22341	37.21%	226	12016	12134	21.16%
2018	80	3792	4073	24.87%	29	6943	7265	33.98%	103	14503	22860	42.67%	250	15253	15575	22.72%
019	64	2853	3074	29.64%	17	4325	4654	37.43%	84	8104	14697	44.04%	187	11376	11713	25.61%

Supplementary table 6 Organisms: identifiability as obtained through reanalysis of the dataset of Menke, Roelandse, Ozyurt, Martone and Bandrowski [5]

	BMC				Nature				Other Journals with guidelines				Journals without guidelines			
	total number of journals analyzed	total number of articles analyzed	total number of organisms found	average % identi-fiable per year	total number of journals analyzed	total number of articles analyzed	total number of organisms found	average % identi-fiable per year	total number of journals analyzed	total number of articles analyzed	total number of organisms found	average % identi-fiable per year	total number of journals analyzed	total number of articles analyzed	total number of organisms found	average % identi-fiable per year
1996	0	0	0	0	0	0	0	0	1	39	36	31%	0	0	0	0
1997	0	0	0	0	0	0	0	0	2	911	632	21%	0	0	0	0
1998	0	0	0	0	0	0	0	0	2	987	885	21%	0	0	0	0
1999	0	0	0	0	0	0	0	0	2	815	713	20%	0	0	0	0
2000	0	0	0	0	0	0	0	0	2	848	801	17%	0	0	0	0
2001	5	107	84	31%	0	0	0	0	2	731	804	12%	0	0	0	0
2002	6	137	72	20%	1	499	63	32%	2	686	776	15%	0	0	0	0
2003	8	356	144	17%	1	608	53	30%	3	764	1123	11%	1	113	17	0%
2004	14	586	238	25%	1	610	40	20%	3	819	996	15%	2	198	32	31%
2005	14	1260	390	27%	1	489	40	35%	7	1008	1219	19%	4	1154	148	21%
2006	20	2144	604	21%	1	483	21	24%	13	1548	1485	18%	6	1074	210	24%
2007	33	3059	1202	20%	1	472	38	37%	15	2601	2208	21%	12	1676	405	23%
2008	37	4033	1393	26%	5	624	148	21%	22	4598	3488	21%	17	2225	780	25%
2009	44	5118	1533	22%	17	1250	860	15%	25	6745	5026	15%	25	2805	1095	20%
2010	54	7440	2346	21%	20	1370	1003	18%	47	10312	6844	21%	39	4060	1997	19%
2011	65	8305	2446	21%	22	1765	1291	18%	59	18072	10412	22%	61	5534	2493	19%
2012	63	9472	2620	22%	28	2577	1896	18%	98	31822	16064	20%	84	7978	3816	21%
2013	77	12135	2853	20%	29	4409	3096	21%	110	43802	20380	22%	97	9919	4135	20%
2014	78	14873	3505	21%	31	5900	4310	19%	141	49435	21294	21%	138	15059	5803	22%
2015	81	14648	3863	19%	33	11675	8013	18%	154	50552	21635	21%	149	18116	7420	20%
2016	78	12190	3718	20%	37	20729	12606	17%	159	49992	20879	23%	163	25180	10202	19%
2017	86	15162	3982	21%	40	26142	15800	18%	175	56493	24796	21%	175	32502	11452	21%
2018	88	16410	4234	21%	39	21879	13514	22%	202	68604	29794	23%	181	30452	10762	21%
2019	68	10186	2971	23%	25	14246	7995	18%	163	48090	20665	24%	129	18900	6148	22%

Supplementary references

1. Horbach SPJM, Halffman W. The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. PloS one 2017;12:e0186281.

2. Perianes-Rodriguez A, Waltman L, Van Eck NJ. Constructing bibliometric networks: A comparison between full and fractional counting. Journal of Informetrics 2016;10:1178-95.

3. Fusenig NE, Capes-Davis A, Bianchini F, Sundell S, Lichter P. The need for a worldwide consensus for cell line authentication: experience implementing a mandatory requirement at the International Journal of Cancer. PLoS biology 2017;15:e2001438.

4. Babic Z, Capes-Davis A, Martone ME, Bairoch A, Ozyurt IB, Gillespie TH, Bandrowski AE. Meta-Research: Incidences of problematic cell lines are lower in papers that use RRIDs to identify cell lines. Elife 2019;8:e41676.

5. Menke J, Roelandse M, Ozyurt B, Martone M, Bandrowski A. The Rigor and Transparency Index Quality Metric for Assessing Biological and Medical Science Methods. Iscience 2020;23:101698.