Supplementary data - Methods

a. Article selection

Articles in the Endnote database were initially filtered by the “find duplicates” tool from the software and subsequently classified using both automatic and manual selection. The order of classification was defined as shown in the flow chart: First “reviews & editorials”, then “clinical studies”, then “in vitro/ex vivo/in silico studies”, then “not rodent preclinical models” and finally “not endometriosis” or “not IC/BPS models”.

The category “Reviews & Editorial Articles” included all types of review publications and editorial comments, letters to editor, opinions, comments, responses to comments, erratum, and others. For the automatic search, articles were filtered by journals that contained the term “review” or “opinion”. Secondly, articles were searched by “type of publication” by containing the term “review”, “editorial”, “comment”, “letter”, “erratum”. Then, the following terms were searched to be contained in either the title, abstract, journal or type of work: “review”, “systematic”, “overview” “update”, “pubmed”, “medline”, “embase”, “scopus”, “database”, “editorial”, “letter”, “re:”, “comment”, “opinion”, “highlight”, “summarize”, “recent evidence”.

The category “Clinical studies” included all types of studies that involved human participation and/or clinical guidelines that might not be selected in the review category. For the automatic search, articles were filtered by containing in their “Keywords” the term “clinical study”. Secondly, articles were searched by containing the following terms either in the title or in the abstract: “patient”, “women”, “men” “diagnos*”, “treatment”, “management”, “prospective”, “retrospective”, “subject”, “randomized”, “controlled trial”, “case”, “case report”, “case of”, “consent”, “recruited”, “hospital” “healthy volunteer”, “prevalen*”, “cross-sectional”, “comorbid”, “questionnaire”, “symptom”, “cohort”, “child”, “elderly”, “adolescent”, “management”, “rare”, “phase”, “follow-up”, “surg*”, “health”.

The category “in vitro/ex vivo/in silico studies” included all studies using exclusively cultured cells (primary or cell lines) or tissue, organ explants, or computer models. For the automatic search the following terms were searched to be contained either in the title or in the abstract: “in vitro”, “ex vivo”, “in silico”, “tissue”, “cell”, “cultured”, “stromal”, “primary”, “nerve preparation”, “bladder strips”, “urothelial cell line”, “biops*”, “detrusor muscle”, “explant”, “human”, “porcine”, “specimen”, “epithelium”, “umbrella cells”, “human bladder”, “cadaver”. If a study included in vitro and in vivo assessments in rodents, then this article was not included in this category but in the category of relevant articles.

The category “Not rodent preclinical models” included all studies that were performed in live animals other than rodents. For the automatic search, the following terms were scanned to be contained either in the title or in the abstract: “baboon”, “macaque”, “monkey”, “primate”, “cat”, “dog”, “bovine”, “feline”.

In all steps described, after each automatic search, a manual confirmation of the search results was performed by reading the title and the abstract to confirm true positives and exclude false positives. If necessary and available, the methods section was also analysed. In cases where insufficient information was available, the publication was categorized as not specified (N/S).
Next, the parallel categories “Not ENDO” or “Not IC/BPS models” were created and included all in vivo rodent preclinical research that was not intrinsically the disease models of ENDO or IC/BPS. In these categories, articles were manually sorted and included models of adenomyosis, ovarian cancer, bladder pain, cystitis (not interstitial), and others.

In the remaining relevant articles, an additional step was undertaken to filter doubling of data: all records that were annotated in the EndNote software as “Conference Abstract” and had led to a publication were also excluded from the analysis. The same procedure was done when poster presentations of the same data were published more than once. Finally, remaining articles were categorized as: “ENDO preclinical rodent models” and “IC/BPS preclinical rodent models” and considered for further definition of subcategories.

b. Definition of subcategories

Articles in both groups “ENDO preclinical rodent models” and “IC/BPS preclinical rodent models” were subclassified as follows: in the ENDO group, preclinical model categories were sorted by: “xenograft”, “syngeneic”, “autologous transplantation”, “allogeneic” “others” and “not specified”. In the IC/BPS group, preclinical model categories were sorted by “bladder-centric”, “complex mechanisms” and “psychological & physical stress”. In both cases, these categories and subcategories were defined following state-of-the-art reviews on animal models of ENDO [3] and IC/BPS [1].

ENDO preclinical models were additionally subclassified by method of delivering endometrial ectopic tissue: either surgical implantation or injection of suspended tissue. In case of surgery, location can be either peritoneum, abdominal wall, ovary, colon, among others. In case of injection, this can be either intraperitoneal or subcutaneous. In the same way, IC/BPS preclinical subcategories were additionally sorted by the procedure used to induce the pathology. First, in the bladder-centric subcategory, articles were further classified by: instillation of irritants, altered expression of urothelial targets, or bladder-targeted radiation. Secondly, in the subcategory models with complex mechanisms, articles were further classified by: injection of pseudorabies virus (PRV) in the tail, comorbid disorders such as intracolonic TBNS instillation, autoimmune models (induced by ovalbumin (OA), Substance P, Loxoribine) and chronic pelvic pain models (vulvodynia, constipation and prostatitis). Finally, in the stress-induced models subcategory, articles were further classified by: water avoidance stress (WAS), neonatal maternal separation (NMS) stress, footshock stress (FS), restraint stress (RS), or manipulation of environmental temperature or light.

c. Scoring system for construct validity

We aimed to create a construct validity (disease mechanism) index for the different preclinical model variants for ENDO. We analysed in each publication the methodology used to induce the pathology and classified them in eight model variants. Autologous transplantation (AT) and allogeneic transplantation (AT) had only one model variant: surgical implantation of tissue fragments (AT-S and AL-S). Syngeneic (SY) and xenograft (XE) model variants had three variants each: those with surgical implantation (SY-S; XE-S), and those where endometriotic/uterine tissue was introduced by injection, either intraperitoneal (SY-IP; XE-IP) or subcutaneous (SY-SC; XE-SC).

We created an algorithm from three questions regarding the methodology used to induce the pathology in each model. Questions were formulated to disclose their resemblance to the
disease pathophysiology. Each model variant was subjected to these questions, and the answers were scored with a three-point scale, depending on how closely a model variant reflects the disease mechanism (either “no effect”, with 0 points, “low effect” with 2 points, and “high effect” with 4 points). The following questions were used, and its rationale is detailed below:

1. **Model resembles endometriotic lesions:** Intraperitoneal injection of suspended tissue resembles very accurately what would occur during retrograde menstruation in patients. Additionally, it provides a considerably less invasive way of delivering endometriotic tissue without causing extensive external or internal tissue injury, compared to surgical techniques. Therefore, we consider that this model variant has a better construct validity than suture techniques. In contrast, subcutaneous lesions do not occur in women, therefore we do not recommend using this technique. For this question, 4 points were given to methodologies using intraperitoneal injection (SY-IP, XE-IP), 2 points to models that performed surgical implantation (AT-S; AL-S; XE-S; SY-S), and 0 points were given to models that performed subcutaneous injections (SY-SC; XE-SC).

2. **Model allows evaluation of the complete immunomodulatory system:** This question intends to clarify the advantages and disadvantages for translational research on the use of uterine vs. endometriotic tissue. While xenograft models have the advantage of using human endometriotic tissue or cells that bring all the disease factors that are actually present in patients, the disadvantage of this model is that for mice to be able to receive tissue of human origin, only immunosuppressed animals can be used. Since ENDO is considered to be an (oestrogen-dependent) inflammatory disorder, the contribution of immunomodulatory components of the recipient mice in this model can only be partially studied. For this question, 4 points were given only to SY-IP, 2 points were given to AT-S, AL-S and SY-S, and 0 points were given to models using xenograft mice (XE-S and XE-IP).

3. **Model allows evaluation of function of gene products by genetic modification:** This question addresses the capability of the model to be used to investigate the role of specific factors by using gene knockout/knockdown or other genetic modifications. For this question, 4 points were given to methodologies using intraperitoneal injection to deliver endometriotic/uterine tissue (SY-IP, XE-IP), 2 points to models that performed surgical implantation of endometriotic/uterine tissue fragments performed in mice (AL-S; XE-S; SY-S), and 0 points were given to the AT-S models variant, since in our search, for this model category no reports using transgenic rats were found.

For IC/BPS, construct validity cannot be scored as the aetiology of the disease is largely unclear.

d. **Analysis of endpoints for Face Validity**

All relevant articles were further classified in terms of the endpoints measured in them to reflect disease symptoms (face validity). For ENDO, endpoints were: “lesion size”, “inflammation”, “adhesion”, “sub-fertility”, as well as “nociception/pain”, and within the latter, articles were further classified by “evoked pain”, and/or “non-evoked pain”. For IC/BPS, endpoints were “bladder morphology”, “bladder function”, “nociception/pain” (also “evoked”, and “non-evoked”), as well as “others”.

The endpoint “lesion size” included all means of analysing and quantifying occurrence of lesions: number, weight, volume, and others. The endpoint “adhesion” required performing
a macroscopic observation of the endometriotic tissue and evaluating the number and/or extent of adhesions by following a scoring system (e.g. [2]). The endpoint “inflammation” involved all means of analysing the presence of inflammatory components (e.g. macrophages, mast cells, neutrophils) in the target tissue, including histological tissue staining or immunostaining, ELISA or other cytokine assays, as well as other molecular biology techniques (e.g RT-qPCR, Western Blot and flow cytometry). The endpoint “sub-fertility” included all publications that evaluated fertility outcomes of the animals in which the model was induced. These included the number and/or survival of embryos, number of pups and gestation time.

The endpoint “nociception/pain” included all publications that assessed pain-related behaviours, and these were further classified as evoked or non-evoked, depending on the nature of the measuring system. In cases where pain thresholds were measured and quantified, these articles were classified as “evoked”, while when spontaneous pain behaviour following observation periods was analysed and quantified, these articles were classified as “non-evoked” pain.

The endpoint “bladder morphology” included all means of analysing and/or quantifying bladder tissue: histological tissue staining or immunostaining, macroscopic evaluations, and others. In this case, lesions and inflammation are included in this endpoint, in contrast to the ENDO approach. The endpoint “bladder function” comprised all means of analysing and quantifying bladder activity and included cystometry, voiding spot assays, analyses of micturition patterns in metabolic cages and others. Finally, the endpoint “others” summarized all assays that have evaluated other aspects in the model, like gene expression and/or gene product function in other tissues than the bladder, or non-structural gene expression analyses from bladder tissue (like RNA Seq, microarrays, and others).

e. Data analysis

Results are presented as percentage proportions from total. Data were inserted and figures were created with the software GraphPad Prism 7.04 (GraphPad Software, Inc). Figure 1 was created with the software Adobe Illustrator 24.0.2 (Adobe Inc., USA).

References